

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENG. DE PRODUÇÃO**

**Metodologia para Modelagem do Padrão da Paisagem integrando
Sistema de Informações Geográficas, Sensoriamento Remoto e
Rede Neural**

Tese apresentada como requisito parcial à
obtenção do grau de Doutor em Engenharia
de Produção.

VICENTE FERNANDO SILVEIRA

Orientadora: Profa. Édis Mafra Lapolli, Dra.

FLORIANÓPOLIS, abril de 1999

Metodologia para Modelagem do Padrão da Paisagem integrando
Sistema de Informações Geográficas, Sensoriamento Remoto e
Rede Neural

VICENTE FERNANDO SILVEIRA

Esta tese foi julgada adequada para a obtenção do grau de Doutor em Engenharia de Produção e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção.

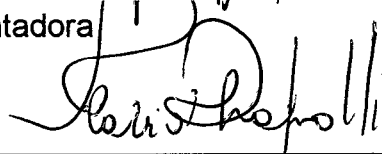


Prof. Ricardo Miranda Barcia, Ph.D
Coordenador do Curso

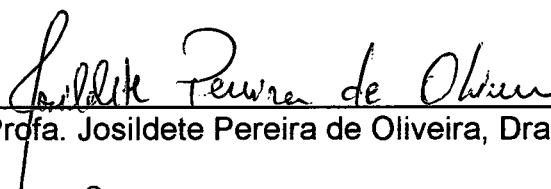
BANCA EXAMINADORA:



Profa. Edis Maíra Lapolli, Dra.
Orientadora



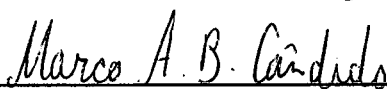
Prof. Flávio Rubens Lapolli, Dr.



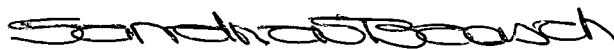
Profa. Josildete Pereira de Oliveira, Dra.



Profa. Lia Caetano Bastos, Dra.



Prof. Marco Antônio Barbosa Cândido, Dr.



Profa. Sandra Sulamita Nahas Baasch, Dra.
Moderadora

We share the same biology, regardless of ideology

Sting

Ao meu pai, Vicente Ferreira Silveira, *in memoriam*, à minha mãe, Myrses De Rosa Silveira e às minhas filhas Sara e Isadora.

Agradecimentos

por ordem aleatória,

ao amigo e professor Marco Antônio Barbosa “Cândido” pela amizade e o incentivo pessoal e profissional,

à professora e orientadora Édis Mafra Lapolli pela orientação e amizade,

à amiga e professora Sandra Sulamita Nahas Baasch pela dedicação constante em todos os aspectos da vida,

ao professor e amigo Ricardo Miranda Barcia por ter possibilitado esse trabalho,

ao colega e amigo, David Vieira da Rosa Fernandes, da Fundação do Meio Ambiente pelo auxílio constante na difícil área da formatação e transferência de dados,

ao bacharel em Ciências da Computação e amigo, “Moreira”, do INPE pelo auxílio na gravação e formatação das imagens,

ao professor Khator, da University of South Florida, pela orientação e incentivo durante o Doutorado em Tampa - Flórida,

aos amigos da Fundação do Meio Ambiente/SC pelos dezoito anos de convívio,

ao colega venezuelano em Tampa, Gerardo Colmenares, pelos “insights” sobre redes neurais, estatística e a participação no desenvolvimento do aplicativo,

à professora Renu, da University of South Florida pelo estímulo no desenvolvimento da tese,

ao CNPq pelo auxílio financeiro em parte da trajetória para conclusão desse trabalho,

aos meus familiares, pelo carinho e compreensão nos momentos mais difíceis da vida,

e à todos, que direta ou indiretamente contribuíram para esse trabalho.

Sumário

| | |
|---|-------|
| Lista de Tabelas..... | VIII |
| Lista de Figuras..... | IX |
| Summary..... | XI |
| Abstract..... | XII |
| | I |
| 1. Introdução..... | 1 |
| 1.1 Origem do Trabalho..... | 1 |
| 1.2 Objetivos do Trabalho..... | 2 |
| 1.3 Justificativa e Importância do Trabalho..... | 4 |
| 1.4 Estrutura do Trabalho..... | 7 |
| 2. Fundamentação Teórica..... | 9 |
| 2.1 Introdução..... | 9 |
| 2.2 Planejamento e Administração Ambiental..... | 9 |
| 2.2.1 Fatores Intervenientes nas Avaliações Ambientais..... | 12 |
| 2.2.2 Avaliação de Impactos Ambientais..... | 16 |
| 2.3 Análise Quantitativa..... | 17 |
| 2.3.1 Dados..... | 17 |
| 2.3.2 Medições e Análises de Dados..... | 19 |
| 2.4 Ecologia da Paisagem..... | 26 |
| 2.4.1 Introdução..... | 26 |
| 2.4.2 Estrutura e Função da Paisagem..... | 28 |
| 2.4.3 Modificação da Paisagem..... | 30 |
| 2.5 Sistemas de Informações Geográficas (GIS)..... | 34 |

| | |
|--|----|
| 2.5.1 Histórico..... | 34 |
| 2.5.2 Elementos Essenciais..... | 36 |
| 2.5.3 Estrutura dos Dados..... | 38 |
| 2.5.4 Capacidades Analíticas..... | 39 |
| 2.5.5 Modelagem Ambiental e GIS..... | 40 |
| 2.6 Sensoriamento Remoto (SR)..... | 45 |
| 2.6.1 Definição de Sensoriamento Remoto..... | 45 |
| 2.6.2 Processamento Digital no SR..... | 48 |
| 2.6.3 SR e GIS..... | 50 |
| 2.7 Redes Neurais Artificiais..... | 53 |
| 2.7.1 Redes Neurais..... | 53 |
| 2.7.2 Modelagem Estatística..... | 56 |
| 2.7.3 Backpropagation..... | 58 |
| 3. Desenvolvimento do Modelo..... | 60 |
| 3.1 Introdução..... | 60 |
| 3.2 Construção do Modelo..... | 62 |
| 3.2.1 Antecedentes..... | 62 |
| 3.2.2 Desenvolvimento..... | 64 |
| 3.3 Representação dos Dados..... | 68 |
| 3.3.1 Variáveis de Entrada..... | 68 |
| 3.3.2 Variável de Saída..... | 70 |
| 3.3.3 Amostragem dos Dados..... | 73 |
| 4. Aplicação Prática..... | 77 |
| 4.1 Caracterização da Área de Estudo..... | 77 |
| 4.1.1 Aspectos Físicos..... | 77 |
| 4.1.2 Situação Ambiental..... | 82 |
| 4.2 A base de dados..... | 84 |

| | |
|--|-----|
| 4.3 Preparação e Processamento dos Dados..... | 85 |
| 4.3.1 Variáveis Independentes..... | 85 |
| 4.3.2 Variável Dependente | 87 |
| 4.3.3 Amostragem e Arquivos de Valor | 93 |
| 4.3.4 Aplicativo para processamento na RNA..... | 98 |
| 4.4 Análise dos Dados..... | 105 |
| 4.5 Considerações Finais..... | 107 |
| | |
| 5. Conclusões e Sugestões para Futuros Trabalhos.... | 112 |
| | |
| 5.1 Conclusões..... | 112 |
| 5.2 Sugestões para futuros trabalhos..... | 115 |
| | |
| Anexos | 116 |
| | |
| 6. Referências..... | 120 |

Lista de Tabelas

| | |
|--|-----|
| Tabela 1 - Comparação entre dados Raster e Vector..... | 41 |
| Tabela 2 - Exemplos de Resolução Espacial..... | 49 |
| Tabela 3 - Operações de manipulação de dados pelos GIS..... | 51 |
| Tabela 4 - Variáveis Sociais..... | 90 |
| Tabela 5 - Variáveis Econômicas..... | 90 |
| Tabela 6 - Variáveis Bio-Físicas..... | 91 |
| Tabela 7 - Coeficiente de Correlação entre as variáveis..... | 108 |

Lista de Figuras

Capítulo 2

| | |
|---|----|
| Fig.1 - Perspectiva do procedimento nos GIS..... | 42 |
| Fig. 2 - Grau crescente de complexidade para os GIS..... | 43 |
| Fig. 3 - Integração entre os GIS e os Modelos Ambientais..... | 45 |
| Fig. 4 - Modelo de Rede Neural de um único Neurônio com uma entrada..... | 54 |
| Fig. 5 - Modelo de Rede Neural de um único Neurônio com várias entradas..... | 55 |
| Fig.6 - Modelo de Rede Neural com vários Neurônios e várias entradas..... | 56 |

Capítulo 3

| | |
|---|----|
| Fig. 7 - Modelo baseado na Integração de SR, GIS e RNA..... | 69 |
| Fig. 8 - Geração das variáveis independentes..... | 71 |
| Fig. 9 - Distância Euclidiana no espaço..... | 74 |
| Fig. 10 - Procedimento de Amostragem..... | 76 |

Capítulo 4

| | |
|--|----|
| Fig. 11 - Plano de Informação Vetorial com as divisões políticas.. | 88 |
|--|----|

| | |
|---|-----|
| Fig. 12 - Tabela do Banco de Dados relacional..... | 89 |
| Fig. 13 - Modelo de Elevação Digital (DEM) gerado por interpolação a partir de isolinhas hipsométricas..... | 92 |
| Fig. 14 - Área da Região do rio Cubatão - Joinville/SC..... | 94 |
| Fig. 15 - Imagem resultante (distance.img) da Análise do Vetor de Mudança ilustrando a magnitude da mudança..... | 95 |
| Fig. 16 - Histograma da imagem distance.img..... | 96 |
| Fig. 17 - Pontos Vetoriais de amostragem sobre uma imagem de variável de interesse..... | 97 |
| Fig. 18 - Tela inicial com as funções principais..... | 100 |
| Fig. 19 - Tela com as funções de preparação dos dados..... | 101 |
| Fig. 20 - Processamento da rede e modificação de parâmetros.. | 102 |
| Fig. 21 - Funções auxiliares dentro do processamento dos dados..... | 103 |
| Fig. 22 - Resultados de utilização da função de teste da rede..... | 104 |
| Fig. 23 - Redução da SSE (amarelo), aumento da taxa de aprendizagem (magenta)..... | 106 |

RESUMO

A grande maioria das agências ambientais governamentais têm tradicionalmente dividido os esforços de políticas e alocação de recursos humanos e financeiros em dois enfoques básicos: Manejo de Recursos Naturais e Controle da Poluição. Esta situação tem levado, especialmente em países em desenvolvimento, a uma alocação ineficiente de recursos humanos, tecnológicos e financeiros através de seus sistemas nacionais de meio ambiente. Vários problemas, tais como: políticas controversas; estabelecimento de leis equivocadas por antagonismo; aplicação de técnicas não apropriadas; uso de escala de trabalho incorreta; têm sido barreiras para a integração desses campos de ação.

É compreensível que à medida que os sistemas organizados pelo homem em nossa sociedade cresçam mais e mais complexos, o manejo ambiental está se tornando uma matéria cada vez mais difícil de se lidar. Em função da enorme quantidade de informação disponível e os diferentes interesses envolvidos, a necessidade para o melhoramento e a implementação de novos métodos para auxílio à tomada de decisão é extremamente urgente.

Um Sistema de Manejo da Informação é proposto neste trabalho, onde dados ambientais podem ser armazenados, manuseados e analisados. A estrutura do sistema está centralizada em um Sistema de Informações Geográficas (GIS) que pode ser descrito como um poderoso conjunto de ferramentas para a coleta, armazenagem, recuperação, transformação e apresentação de dados espaciais do mundo real para um conjunto particular de objetivos. Os dados de entrada para o modelo vem do Sensoriamento Remoto (imagens de satélite), fotografias aéreas, levantamentos de campo e censos.

Através da geração de mapeamentos derivados pelo GIS uma condição ambiental pode ser representada em qualquer escala temporal ou espacial. Através de uma série temporal comparativa e geração de mapeamentos derivativos dados podem ser extraídos para o treinamento de

uma rede neural artificial (backpropagation) para a previsão de mudanças ambientais.

Dados ambientais multivariados podem ser satisfatoriamente melhor processados através de uma rede neural artificial. Não existe necessidade de conhecer a distribuição estatística dos dados, a rede pode identificar quanto peso cada fonte de dado possui no processo sob análise. As redes neurais artificiais são mais robustas do que os métodos estatísticos convencionais pelo fato de tolerarem imprecisões maiores nos dados (“noise” e “missing data”); possuem adaptação através de uma escala espacial; e são computacionalmente mais eficientes.

Prever mudanças ambientais é um problema metodológico que integra muitos campos de conhecimento específicos. As diferentes escalas temporais e espaciais dos processos ambientais adiciona ainda mais complexidade aos estudos de investigação.

A integração do Sensoriamento Remoto, Sistema de Informações Geográficas e Redes Neurais Artificiais é uma promessa para responder, em parte, estas dificuldades.

Este trabalho investiga essa integração e mostra os resultados obtidos através do modelo proposto.

Abstract

Predicting environmental change give policy makers indications for a better allocation of natural, economic and social resources. Intensive use of these resources may lead to unpredictable results due to the highly non-deterministic characteristics of environmental data. An integrated model using Remote Sensing, Geographic Information System and Artificial Neural Network is proposed in order to assess possible causal relationships that lead to environmental modifications.

1. Introdução

1.1 Origem do Trabalho

O campo de estudo das Ciências Ambientais tem sido usualmente dividido, especialmente em termos de enfoque político-administrativo, entre as áreas do conhecimento científico de Manejo e Conservação de Recursos Naturais e a do Controle da Poluição.

Esta tendência, provavelmente, originou-se no universo acadêmico onde a pesquisa e desenvolvimento do Controle de Poluição é geralmente implementada no campo de Engenharia e Ciências associadas, enquanto que o campo do Manejo e Conservação de Recursos Naturais pertence tradicionalmente à área das Ciências Ambientais. A dicotomia sempre presente entre ciência básica e aplicada também contribuiu para o incremento desta tendência no âmbito deste enfoque ambiental.

Esta situação conduziu, especialmente nos países em desenvolvimento, a uma distribuição ineficiente de recursos humanos, tecnológicos e financeiros, ao longo do tempo, nos sistemas de políticas ambientais nacionais.

Vários problemas persistem relacionados à políticas controversas; execução de leis influenciadas por antagonismo; técnicas diferentes aplicadas; escala de trabalho e outros. Todos esses processos têm sido uma barreira teórica e metodológica para integração desses dois campos de conhecimento.

O reconhecimento público da necessidade de preservação ambiental concomitante à adoção de novas políticas industriais conduziram mais recentemente, nas décadas de oitenta e noventa, para uma convergência e talvez futura integração destes dois campos, através da implementação da disciplina da Gestão Ambiental.

O conhecimento aumentado em processos ecológicos e a necessidade por desenvolver mais eficiência e qualidade na indústria permitiu esta aproximação. Políticas ambientais e processos tecnológicos

devem agora levar em consideração, como um todo, o meio ambiente e o desenvolvimento da sociedade humana.

O novo conceito de linha de produção das fábricas deve considerar as novas implicações sociais e econômicas e estender-se desde os processos de extração e uso dos recursos naturais até a obtenção do produto final e os consequentes resíduos gerados na sua fabricação industrial.

É compreensível que à medida que os sistemas organizados pelo homem na nossa sociedade tornam-se mais e mais complexos, o planejamento e administração dessa sociedade – Gestão Ambiental – torna-se um assunto também de extrema dificuldade para sua compreensão, elaboração e aplicabilidade.

Com o advento da revolução digital uma quantidade impressionante de dados ambientais tornou-se disponível com uma frequência regular e a custos razoáveis. Por outro lado, as diversas correntes ideológicas para a aplicação de políticas de controle ambiental tem ocasionado uma estagnação de maiores progressos para a gestão do meio ambiente.

O esforço de pesquisa e desenvolvimento necessário deve centrar-se na busca da integração desses dados ambientais, sua transformação em informação utilizável e posterior uso, para auxílio à tomada de decisões, nas políticas públicas e privadas de gestão ambiental.

Esforços para integração destes diferentes enfoques ambientais interdisciplinares têm sido feitos, especialmente na comunidade científica. À medida que mais informação fica disponível livremente, as possibilidades de aplicação desses novos conhecimentos na vida real tornam-se razoavelmente plausíveis.

1.2 Objetivos do Trabalho

Objetivo Geral

- Desenvolver um modelo de reconhecimento de associações e descoberta de regularidades dos dados dentro do conjunto de padrões presentes na paisagem, representadas por imagens de satélite, através da

integração entre o potencial de geração de dados ambientais do Sensoriamento Remoto, com o potencial de manipulação, transformação e análise de dados ambientais dos Sistemas de Informações Geográficas, e com o potencial de captura de relacionamentos da Rede Neural Artificial.

Objetivos Específicos

- Contribuir para o desenvolvimento de noções teórico-espaciais novas em ciências ambientais;

- Contribuir para o desenvolvimento de novos enfoques de pesquisa em análise ambiental que gerem conceitos e perspectivas ambientais não convencionais.

- Desenvolver um método para representação e descoberta de associações entre dados complexos, multivariados e de fontes diversas enfatizados em padrões locais;

- Desenvolver um procedimento para manipulação e amostragem de dados espaciais para armazenamento, análise e modelagem dos mesmos;

- Desenvolver um procedimento para geração, manipulação, amostragem e migração de variáveis espaciais de um ambiente de Sistemas de Informações Geográficas para uma Rede Neural Artificial;

- Desenvolver um programa em ambiente matemático (MATLAB® for Windows) que importa imagens raster e arquivos de valor (ASCII) como vetores de entrada - saída estabelece a arquitetura, normaliza os dados, treina e testa uma Rede Neural Artificial;

- Desenvolver um modelo de avaliação de mudanças na paisagem no conjunto de variáveis independentes indutoras.

1.3 Justificativa e Importância do Trabalho

Os problemas

O resumo da aplicação nos últimos 30 anos de análise quantitativa e modelagem matemática de bases de dados espaciais, especialmente quando se considera dados ambientais, contribuiu muito pouco para o desenvolvimento de conceitos teóricos, que estabelecessem novos paradigmas para as Ciências Ambientais (Openshaw, 1994).

Forman (1995) propõe doze princípios gerais de ecologia da paisagem e regional para estimular hipóteses de discussão e pesquisa. No final, ele conclui que após uma década de trabalho em Ecologia da Paisagem ainda existe um processo de solidificação teórica acontecendo, com a finalidade de determinar conceitos fundamentais, através da identificação de princípios gerais emergentes.

A maioria do progresso em análise espacial têm sido feita com métodos de estatística inferencial dos anos sessenta e setenta. Mas, limitações destas técnicas atrasaram um desenvolvimento adicional e interesse pela área que pudessem estabelecer novos conceitos e uma consequente base científica para a continuidade de pesquisa.

Sistemas e processos ambientais contêm uma grande quantidade de componentes estocásticos tornando os modelos de difícil especificação assim como para com padrões e processos presentes dentro destes sistemas. A maioria dos modelos ambientais desenvolvidos são de excelente valor acadêmico mas de aplicação empírica muito improvável. Eles são aperfeiçoados para representação teórica em lugar de representação de dados (Openshaw, 1993).

Nos anos noventa, para agravar ainda mais esta situação, uma nova revolução quantitativa baseada nos Sistemas de Informação Geográfica (Geographic Information Systems/GIS – seção 2.5) tem produzido uma enorme quantidade de dados espaciais sem prover as ferramentas necessárias para a análise e modelagem destes dados.

Apesar de ser uma ferramenta de excelente qualidade para descrever a realidade espacial, os GIS não contribuíram muito para novas descobertas científicas que gerassem, novos conceitos e perspectivas ambientais. Grande parte dos investigadores têm usado os GIS como um fim em si mesmo ao invés de utilizá-los como uma ferramenta auxiliar para problemas ambientais de conteúdo espacial.

A era corrente dos GIS tem promovido muito mais o armazenamento de dados do que gerar conceitos novos para entender os padrões e processos que podem acontecer no meio ambiente.

Há uma falta profunda de técnicas e conceitos teóricos para examinar padrões prováveis de regularidades presentes em nossas bases de dados espaciais. A maioria das realizações teóricas têm falhado em provar hipóteses de uma maneira consistente, ou ainda validar teorias através da produção de previsões empiricamente viáveis (Openshaw, 1994).

Talvez, parte do problema, possa ser explicado pela existência de uma enorme lacuna entre investigadores que estão baseados em diferentes filosofias e paradigmas de pesquisa. Este fato pode estar causando equívocos teóricos e técnicos quando se procura estabelecer princípios gerais para as Ciências Ambientais.

As transformações

A Ciência da Meteorologia talvez seja o melhor exemplo, de que os sistemas determinísticos não são suficientes para descrever e fazer previsões de fenômenos ambientais. Há uma tendência atual de que as análises estatísticas e matemáticas podem ser substituídas, em alguns problemas, por ferramentas de Inteligência Artificial (IA – seção 2.7) com muitas vantagens e custo mais baixo, reduzindo tempo e recursos humanos para resolver a complexidade de problemas ambientais.

Nos anos noventa o aumento do desempenho e a diminuição do preço dos programas e equipamentos digitais estimulou, entre outros fatores, o desenvolvimento de tecnologias de inteligência artificial.

Conforme comentado anteriormente, vastas quantidades de informação espacial e dados ambientais relevantes têm sido gerados pelos GIS. A maioria da informação espacial sobre a Terra tem se tornado disponível em bases de dados, algumas delas até mesmo em endereços da Internet (USA Geological Survey, NASA, EPA).

Os GIS têm promovido um enfoque integrado para os problemas de manipulação de dados nas mais diversas áreas das Ciências e em aplicações industriais e comerciais. Openshaw (1994a) considera que a Análise Espacial passou por três revoluções quantitativas: nos anos sessenta ela foi estatística, nos setenta de modelagem matemática e agora, nos noventa, a de neurocomputação.

A aplicabilidade de técnicas de IA são quase irrestritas e sua aplicação nas Ciências Ambientais começou praticamente nesta década. Alguns dos problemas, conforme já mencionado algumas vezes no texto, gerados por esta nova base de dados disponível podem ser resolvidos por programas e equipamentos digitais emergentes (especialmente para captura, manipulação, transformação e armazenamento) e alguns por novas técnicas de análise e modelagem (IA).

As soluções

As bases de dados providas por fontes diversas têm dados complexos e multivariados. Para o entendimento da estrutura e regularidades dos padrões destes dados resumos descritivos têm sido produzidos. Métodos quantitativos estatísticos têm conseguido, de alguma maneira, lidar com esta tarefa.

Entretanto com o aumento sempre constante dessa base de dados, o significado desta redução de dados é cada vez menos fidedigno, sendo incapaz de demonstrar as possíveis funções e padrões dos processos de interesse.

Talvez uma nova direção de pesquisa seja em novas ferramentas baseada em computadores avançados (paralelos) explorando mais a

estrutura dos dados (empírica) do que a teoria. Os modelos novos terão que ser criados mais a partir dos dados do que da teoria (Openshaw, 1993).

A seleção dos dados parece ser agora a decisão mais pertinente. A escolha de decidir quais variáveis podem ser relevantes, semi-relevantes ou totalmente irrelevantes deveria ser feita pela informação presente nos próprios dados. Os métodos de análise deveriam ser enfatizados nos padrões locais e nos relacionamentos dos fenômenos ambientais ao invés de procurar por relações causais presumivelmente explicadas por conceitos determinísticos.

O desafio para a próxima década é aglutinar a imensa capacidade de manuseio de dados dos GIS com técnicas de análise computacional e direcionar este esforço para noções teórico-espaciais novas em ciências ambientais.

As novas tecnologias de IA têm uma aplicabilidade extensiva e este fato pode conduzir a uma complementação progressiva na modelagem matemática e técnicas estatísticas em áreas de conhecimento aonde estas técnicas não demonstraram ser o bastante para gerar percepções e conceitos novos.

Considerando suas imensas potencialidades cibernéticas os GIS podem ser uma tecnologia de liderança nas próximas décadas porque podem unir as Ciências Ambientais com os conceitos e técnicas de IA.

1.4 Estrutura do Trabalho

O presente trabalho foi estruturado em seis partes, a saber: Introdução (1), Fundamentação Teórica (2), Desenvolvimento do Modelo (3), Aplicação Prática (4), Conclusões e Sugestões para Futuros Trabalhos (5) e Referências (6).

Na primeira parte, são delineadas as circunstâncias da origem do trabalho, que foi influenciado predominantemente pelas interações teórico-práticas existentes entre os *orgãos gestores* das Políticas Ambientais e o *universo acadêmico*, gerador de Ciência e Tecnologia para Gestão do Meio Ambiente. A justificativa e importância do trabalho também são explanadas

apontando os problemas, as transformações e possíveis soluções para a questão da análise ambiental, com especial ênfase nos últimos trinta anos.

Na segunda parte, uma Fundamentação Teórica é demonstrada para estabelecer os conceitos que serão utilizados posteriormente quando da construção do modelo proposto. Inicialmente discute-se os diferentes enfoques para o planejamento e administração ambiental (Gestão Ambiental), considerando os diversos fatores intervenientes na análise ambiental, cujos conceitos de Ecologia da Paisagem também são fundamentais para essa conceituação teórica. Em seguida uma retrospectiva dos principais métodos quantitativos que têm relação direta com o desenvolvimento do presente trabalho é apresentada.

Finalmente, nessa segunda parte, a trilogia fundamental do modelo proposto é apresentada e discutida: os Sistemas de Informações Geográficas, o Sensoriamento Remoto e as Redes Neurais Artificiais.

Na terceira parte, as fases de elaboração do modelo proposto são discutidas. Para a construção do modelo, o arcabouço teórico que permite a sua fundamentação é discutido. Em seguida, a maneira como os dados podem ser representados dentro desse modelo são demonstrados.

Na quarta parte, a aplicação do desenvolvimento teórico é efetuada. Preliminarmente uma caracterização da área de estudo é demonstrada. Em seguida, a base de dados utilizada, analógica e digital, é apresentada para que na próxima fase os dados possam ser preparados e processados pelo modelo. Enfim, os dados são analisados e os parâmetros utilizados para os diversos ensaios do modelo, por exemplo, modificação do conjunto de variáveis independentes; definição da arquitetura geral da rede; sementeira inicial com números aleatórios e produção de uma saída desejada, são mostrados.

Na quinta parte os resultados e conclusões são discutidos e, a partir das constatações obtidas, são elaboradas sugestões para futuros trabalhos.

Finalmente, na sexta e última parte, a lista de referências bibliográficas que embasaram o presente trabalho é descrita.

2. Fundamentação Teórica

2.1 Introdução

A conceituação do meio ambiente é complexa, multidisciplinar e não-determinística. Um único campo de conhecimento da cultura humana não é o suficiente para definir e estabelecer paradigmas que possam explicar a ocorrência de todos os fenômenos ambientais possíveis.

A Ecologia, provavelmente a disciplina mais antiga a estudar o meio ambiente, passou, nas últimas décadas, a incorporar a multidisciplinaridade necessária que os estudos ambientais requerem.

A Engenharia, a Economia, a Medicina, o Direito, entre outros, são exemplos de campos de conhecimento que passaram a se preocupar com a problemática ambiental. Cada um com as suas tendências ("bias") próprias, suas preocupações e expectativas.

Nesta última década, um esforço descomunal de desenvolvimento teórico e prático na área ambiental tem sido feito. Os trabalhos desenvolvidos partem de escala microscópica para modelos globais de interações ambientais.

A urgência de solução dos problemas ambientais requer uma flexibilidade de aplicação dos conceitos tradicionalmente defendidos no meio acadêmico, paralelo à execução de políticas ambientais mais efetivas e comprometidas, alicerçadas, mesmo que paradoxalmente, à um sólido conhecimento técnico-científico.

As seções seguintes, dentro do capítulo Fundamentação Teórica, discorrem sobre algumas disciplinas que direta ou indiretamente têm aplicação e contribuem para as Ciências Ambientais.

2.2 Planejamento e Administração Ambiental

Determinar as relações de causa e efeito entre o ser humano e o meio ambiente tem uma tendência muito forte de enveredar por uma vertente ideológica em detrimento de uma análise científica.

A influência recíproca homem-natureza acontece desde os tempos imemoriais e diagnósticos ambientais mais recentes evidenciam a existência de uma forte correlação entre o tipo de desenvolvimento da sociedade humana e graus diversos de degradação ambiental.

Este fato parece sugerir que talvez os esforços de pesquisas devessem centrar-se mais no desenvolvimento de métodos de prognósticos para a prevenção de modificações ambientais de efeitos danosos, ao invés de simples inventários indicativos de uma situação ambiental momentânea.

Neste contexto uma definição de Planejamento e Administração Ambiental – Gestão Ambiental, tendo em vista a amplitude do assunto, torna-se um tema de difícil conceituação, sob alguns aspectos até um pouco vago.

Uma das questões essenciais para qualquer tipo de avaliação ambiental é o desenvolvimento metodológico da análise dos vários fatores que compõem e influenciam o fenômeno que se procura entender, explicar e prever. Em outras palavras, quando possíveis efeitos, independentemente se benéficos ou prejudiciais, de modificações ambientais são avaliados, a limitação de instrumentos e métodos de análise deve ser fortemente considerada.

Outro ponto a considerar é a tendência que cada profissional tem de basear seu enfoque de pesquisa em paradigmas e métodos que pertencem ao seu campo de estudo. A pesquisa ambiental conduz a vários dilemas, cujos fatores de ordem social, econômica ou política interagem com aqueles de ordem natural. Talvez como consequência do procedimento habitual de avaliação científica, essa tendência apareça frequentemente, causando uma fragmentação do problema.

Muito foi dito sobre a importância de um enfoque multidisciplinar para os estudos ambientais. Porém, as restrições metodológicas para integração das diferentes disciplinas envolvidas têm sido um fator limitante para o estabelecimento de políticas ambientais convergentes.

Benediktsson et al. (1990) argumenta que nas últimas décadas, progressos em tecnologias espacial e digital tornaram possível a acumulação

de quantidades enormes de dados relacionadas com o meio ambiente. O mesmo lugar de interesse, pode conter muitos tipos de dados de fontes diferentes ("multisource data"). Apesar de promover novos desafios de pesquisa este fato pode comprometer os esforços para uma gestão ambiental adequada tendo em vista a profusão de dados disponíveis mas não transformados em informação.

Seguramente, um dos grandes desafios sobre a análise ambiental para o futuro é a integração dos dados já disponíveis. A partir deles, pode-se obter uma qualidade maior das informações, tornando os prognósticos e diagnósticos mais precisos e factíveis.

Nas últimas décadas, a problemática ambiental era considerada ingênua e despropositada do ponto de vista dos políticos. Mais recentemente, ela passou da posição de ser considerada apenas uma externalidade pelos economistas, para fazer parte obrigatória de qualquer agenda de desenvolvimento econômico-social sensata.

É impossível desconsiderar a importância estratégica para a soberania nacional que o conhecimento e o manejo correto dos Recursos Naturais têm.

A utilização desmesurada do petróleo; dos potenciais de água doce; da biomassa fotossintética natural e plantada; dos minérios; da fertilidade do solo; das populações animais marinhas; do poder tampão físico-químico da atmosfera; e, por que não? Dos recursos humanos e financeiros dos países menos favorecidos ou emergentes pelos países ricos, mostram-se como políticas de desenvolvimento econômico e social não sustentáveis a médio e longo prazos.

O que parece ingênuo ("naïve") agora, nos anos noventa, é a implementação de Políticas de Produção Industriais e Agropecuárias que desconsiderem planos de integração e controle ambiental. Em essência, não é somente a necessidade para o estabelecimento de procedimentos para análises ambientais com rigor científico está em questão, mas o que se tem em mente também é a aplicabilidade dessas novas propostas considerando o caráter de urgência social e econômica que os problemas ambientais demandam.

2.2.1 Fatores Intervenientes nas Avaliações Ambientais

Variabilidade

Os fenômenos ambientais são difíceis de avaliar e prever tendo em vista sua propagação espacial e temporal aliada a possíveis relações causais entre os diferentes fatores considerados. Correlações na natureza são de difícil explicação devido a alta variabilidade dos fenômenos. A maioria deles é influenciada por fatores causais não-controlados na sua variação e frequentemente não identificáveis.

A categorização de um fenômeno ecológico como determinístico ou probabilístico pode conter diversos erros de interpretação. Muitos destes fenômenos podem ser previstos sob conceitos estritamente mecanicistas e determinísticos. Outros, talvez por causa de nossa falta de habilidade em controlá-los e a variedade de variáveis causais presentes, podem parecer principalmente probabilísticos.

Johnston (1978), em uma tentativa de sistematização do problema dividiu-o em duas perguntas básicas a serem respondidas: o fenômeno repete-se de um modo semelhante em diferentes locais? É nos lugares que acontecem os inter-relacionamentos homem-ambiente, porém, como nós discutiremos mais adiante, várias dificuldades existem para definir um lugar de investigação de um fenômeno.

Quando consideramos qualquer fenômeno ambiental, obviamente ele não se origina de só uma causa. Muitas variáveis estarão interagindo, tanto separadamente como combinadas. A compreensão de todo o processo envolve um conhecimento individual das causas, havendo a necessidade de uma amostragem cuidadosa e o uso de evidências.

Em um trabalho empírico isso envolve a observação de eventos no mundo real, nos lugares. Entretanto, relações de causa-efeito são geralmente difíceis de isolar, até mesmo quando podem ser descritas associações entre as variáveis pertinentes. Os relacionamentos são geralmente mais descritivos (associados) do que propriamente causais.

A segunda pergunta que o autor propõe é: o lugar influencia ou determina o fenômeno? Lugares diferentes só podem ser descritos como diferentes em termos de enfoques de análise semelhantes. Por exemplo, uma cartografia comparativa de um lugar com uma série de outros lugares. Só podem ser entendidos processos e padrões únicos em termos de conceitos gerais de inter-relação entre variáveis, explicadas em um conceito específico para aquele local.

Em outras palavras, pode-se considerar que alguns mecanismos de inter-relação entre determinadas variáveis tenham um padrão estabelecido e sejam razoavelmente interpretáveis. Entretanto estes mecanismos não se comportariam da mesma maneira quando fossem considerados em locais diferentes.

Escala

Apesar das restrições que a variabilidade natural dos fenômenos impõe, vários enfoques, relativos à estudos ambientais, podem ser levados a cabo.

Em geral, para a delimitação do problema, uma definição da escala espacial e temporal de trabalho mostra-se metodologicamente sensata para a maior circunscrição possível da variabilidade e dos limites de ação do fenômeno.

Turner et al. (1994) descrevem métodos disponíveis de "detecção de escala" para utilização de técnicas apropriadas para resolver problemas específicos.

Lee et al. (1992) indicam que a área de interesse para estudo é delimitada pelos elementos espaciais e processos sob consideração. A mudança de escalas mudaria a relevância das variáveis pertinentes, até mesmo processos causais podem diferir quando enfocados sob escalas diferentes.

Meentemeyer (1989) argumenta que a falta de controle experimental e o tamanho da unidade de observação são os dois maiores problemas nas análises ambientais.

Para uma utilização correta dos resultados de qualquer pesquisa, eles devem ser gerais e válidos em outros lugares. Desta maneira o controle da escala é uma premissa para o delineamento do problema. Atribuindo escala a uma análise quantitativa para um estudo ambiental é uma maneira de tornar um princípio relativo operacional (Forman, 1995).

Wiens (1992) em um levantamento sobre os tópicos publicados no periódico *Landscape Ecology* mostra que problemas relacionados com escala ocuparam 24% do foco de atenção dos artigos. Ele também relatou que os estudos foram conduzidos principalmente (73%) em grandes áreas (ha ou Km).

A maioria das variáveis biofísicas ambientais são extremamente complexas e, à medida que variações de resolução espacial ocorrem, a informação que determinada cena contém pode tornar-se infinita. Adicione-se a esse fato a interface sócio-econômica e muito poucas suposições lógicas podem ser inferidas sobre o assunto. Este tipo de pensamento pode conduzir ao que Meentemeyer (1989) concluiu que "no fim, parece que as escalas de trabalho são selecionadas inconscientemente e parecem então serem completamente arbitrárias".

Limites

O estabelecimento dos limites de um modelo é uma necessidade para delimitar o universo de interesse. Isto pode ser considerado como uma tentativa de tornar o mundo real discreto ("discretization"), o estabelecimento de uma metodologia que possa ser aplicada em qualquer situação com estrutura semelhante.

A bacia hidrográfica - uma região geográfica que drena água para um rio, um sistema de rios, ou outro corpo de água - tem sido usada para avaliar vários assuntos que dizem respeito à gestão ambiental. Os ecossistemas aquáticos e sua relação com os sistemas terrestres podem ser melhor entendidos pelo fato de tal enfoque poder confinar a maioria dos fatores relevantes (bióticos e abióticos) que possam induzir mudanças dentro daquele sistema, por suposição, fechado.

Este limite parece naturalmente mais lógico quando nós comparamos, por exemplo, com divisões políticas (distrito, município, regiões, estado). O elemento água pode ser interpretado como um indicador dinâmico dentro deste sistema, agindo como uma testemunha da situação atual e passada do seu entorno. Estudos tradicionais tomam a bacia hidrográfica como a unidade de estudo e várias medidas são efetuadas nesta escala.

O enfoque da bacia hidrográfica não pode limitar o problema como um todo. Um perímetro de uma bacia hidrográfica pode ter de alguns quilômetros quadrados (normalmente classificados como sub-bacias) para milhares e às vezes milhões de quilômetros quadrados (Rio Amazonas).

Decisões sobre amplitude de pesquisa são imprescindíveis para iniciar qualquer estudo ambiental que vise monitorar, analisar e diagnosticar. Além disso vários fenômenos, tanto ecológicos como sociais, extrapolam os limites impostos por esta aproximação.

Escolha de Variáveis

Assim como existe um miríade de fontes de dados, medições incontáveis podem ser feitas, só limitadas por nossa imaginação (Kachigan, 1982).

As características das análises a serem executadas vão se basear principalmente na formação do analista e nas questões que ele deseja responder (Turner et al. 1991).

Desta maneira, o que supostamente pode definir melhor o problema é o processo de escolha das variáveis, os possíveis inter-relacionamentos entre elas e qual técnica é mais apropriada para a análise destas variáveis.

O processo de escolha das variáveis de análise pode ser uma ferramenta extremamente efetiva para pelo menos reduzir os problemas descritos previamente. Com a redução do número de variáveis os problemas de variabilidade são minimizados. O tipo de variável escolhido também define a escala de trabalho e pode delimitar o alcance da área de estudo.

Porém uma compensação deve ser pesquisada, de forma que o número e tipo de variáveis escolhidas possam produzir confiabilidade e

significado, ou seja, a decisão para esta escolha deveria passar por um método de tentativa e erro de tipo. Fazendo isto, a massa de dados pode ela mesmo indicar padrões estatísticos com significado para análise, ao invés da escolha no processo ser racionalmente humana.

Como discutido no princípio desta seção, o número de possíveis variáveis ambientais a ser escolhido é infinito o que faz o entendimento de possíveis relacionamentos causais uma tarefa quase irrealizável.

Portanto uma divisão entre variáveis sócio-econômicas e biofísicas pode ser usada para demarcar um espaço de discrepância para simplificação do modelo. Um levantamento no universo de todas as áreas das ciências que tenham relação com a questão ambiental, mostra uma distinção natural entre o que poderia ser definido como cultural - feito pelo homem - e como biofísico - feito pela natureza.

2.2.2 Avaliação de Impactos Ambientais

A Avaliação de Impacto Ambiental constitui instrumento de real importância e atualidade para a preservação e o manejo do meio ambiente.

Qualquer alteração nas propriedades físicas, químicas e biológicas do meio ambiente causada por atividade humana que afetem a saúde e o bem-estar da população; as atividades sociais e econômicas; os ecossistemas, pode ser considerada como impacto ambiental.

Nesse contexto, o estudo de impacto ambiental deve ser um instrumento de prevenção e controlador da degradação ambiental. Trata-se de uma técnica de identificação de risco e de informação antecipada, objetivando avaliar os impactos sobre o meio ambiente de quaisquer ações ou omissões que possam ocorrer decorrentes de atividades humanas.

O estudo de impacto ambiental precede à avaliação ambiental e constitui-se de um relatório dos riscos prováveis ao meio ambiente decorrentes da atividade projetada. O relatório é submetido à avaliação ambiental para a escolha ou não, das alternativas de desenvolvimento da atividade proposta.

No Brasil, quando se analisa o ordenamento jurídico, o estudo de impacto ambiental e a avaliação posterior está implicitamente presente em várias normas, a saber:

- O Estatuto da Terra (...o uso racional do solo e à conservação dos recursos naturais, em face de sua função social...);
- O Código Florestal (... as florestas e demais formas de vegetação são bens de interesse comum a todos os habitantes do país...);
- O Decreto-Lei no. 1.413, de 14/08/75 (...dispõe sobre o controle da poluição do meio ambiente...).

O Estudo de Impacto Ambiental foi introduzido pela Lei no. 6.803, de 02/07/80 que determina que, a aprovação de zonas de uso estritamente industrial será precedida de estudos especiais de alternativas e avaliações de impacto, visando minimizar os riscos da solução adotada. Posteriormente foi adotado de forma extensiva a todas as áreas passíveis de degradação ambiental pela Lei no. 6.938, de 31/08/81 (Política Nacional de Meio Ambiente) e a Resolução 001/86.

Entretanto, apesar de todos os avanços jurídicos apresentados resumidamente aqui, é notória a não aplicação das normas legais e regulamentares agravando ainda mais a deterioração do nosso patrimônio ambiental.

Uma gestão adequada do meio ambiente requer uma visão multidisciplinar e inter-institucional. Numa possível hierarquização dos fatores pertinentes a essa questão, a estrutura e conjuntura política e jurídica, alicerçadas na política econômica e industrial, são preponderantes.

A pressão social tem o aspecto reativo, quando as ações políticas prolongam-se por muito tempo como inadequadas. Cabe à comunidade acadêmica contribuir para o bem-estar social provendo os meios educacionais e técnicos de desenvolvimento e aplicabilidade da Gestão Ambiental.

2.3 Análise Quantitativa

2.3.1 Dados

Observações de algumas características ou propriedades dos objetos (normalmente as pessoas ou coisas), medidas na menor unidade de amostragem, define o conceito de *dados*.

Estes objetos ou as pessoas nas quais as medidas (ou observações) são feitas podem ser chamados de *unidades experimentais*. A natureza dos dados ou as propriedades em um estudo, medida pelas observações individuais é a *variável*.

Observações medidas em uma escala numérica são definidas como dados *quantitativos* enquanto que dados não numéricos só podem ser classificados em um grupo de categorias e são denominados dados *qualitativos*.

Existe uma grande gama de representação numérica - *escalas* - dos valores de uma variável (tipos de dados), entretanto, via de regra, podem ser definidas quatro categorias gerais:

Nominal – este nível de medida não têm nenhuma estimacão numérica e envolve só categorização. Os objetos têm características que diferem somente em tipo. Não há algum tipo de julgamento comparativo. Ela só pode representar valores diferentes de uma determinada variável. A forma mais básica é a representação binária, ou presença-ausência. Na maioria das vezes esse tipo de medida é referida como *variável qualitativa*.

Ordinal – os dados podem ser colocados em posição de grau hierárquico em uma ordem de significância, ao longo de um contínuo. Apesar de possuir a propriedade de transitividade somente pode indicar que os números na escala não são valores idênticos. Não há alguma possibilidade para fazer inferências sobre os graus de diferença entre valores na escala, em outras palavras, os números nesta escala não indicam uma relação de um com o outro mas somente indicam que eles não são valores idênticos. Como o nominal, esse tipo de medida também é referida como *variável qualitativa*. Apesar de ser quantitativamente comparáveis estas medidas não teriam algum significado entre variáveis diferentes.

Intervalo – nesta escala intervalos iguais entre os valores da escala tem significado igual. Cada observação tem um valor separado no fenômeno que está sendo medido e as diferenças entre observações adjacentes na escala podem variar.

Uma limitação desta escala é que razões de valores da mesma não têm significado algum. Isto é devido ao fato que tal escala tem um ponto zero arbitrário que não necessariamente representa uma quantidade zero. Por conseguinte, os resultados medidos nesta escala não podem ser interpretados em um sentido absoluto.

Razão – este é o nível mais preciso de medição, contendo mais informação que todas as outras escalas. Difere da escala anterior em um só respeito - a presença de um ponto zero genuíno. Tanto a escala de intervalos como a de razão são consideradas escalas métricas e as variáveis medidas por elas *variáveis quantitativas*.

A maioria das escalas de razão tem unidades para distinguir uma da outra dependendo da praticidade e conveniência. Duas escalas muito comuns que não têm unidades de medida são a escala de frequência e a escala de porcentagem. Elas podem ser consideradas puras no sentido que não requerem qualquer unidade, tendo um significado universal.

Uma *população* é uma coleção ou conjunto de dados que descreve algum fenômeno de interesse. Essa coleção é definida através de uma declaração do que pertence e do que não pertence ao conjunto de dados.

Claro que muitas das populações são muito grandes para medir cada observação individual, assim o que se faz é selecionar um subconjunto de valores de uma população, chamado *amostra*.

A amostra é representada por uma coleção de observações individuais ou obtidas de acordo com um procedimento especificado.

2.3.2 Medições e Análise de Dados

Análises quantitativas podem ser realizadas tanto em termos descritivos como em termos inferenciais.

No primeiro – *análise descritiva* – a preocupação maior é por uma medição exaustiva direta das características da população. Uma medida sumária numérica chamada *parâmetro* descreve uma característica de uma população. Devem ser feitas observações em cada indivíduo da população para obtenção do valor do parâmetro.

Em oposição a isso, a *análise inferencial* mede as características de dados amostrados e então faz conclusões sobre o valor correspondente na população da qual a amostra foi tirada.

Existem várias medidas descritivas numéricas que fazem sumarização dos dados – um resumo de grandes conjuntos de observações em formas mais compactas e interpretáveis – tais como, medidas de *tendência central* (média aritmética, mediana, moda); medidas de variabilidade (alcance, variância, desvio padrão) que nos permitem construir uma imagem mental da frequência relativa da distribuição de um conjunto de dados. O primeiro grupo localiza o centro de uma distribuição de frequência relativa enquanto que o segundo mede a expansão de uma distribuição de frequência relativa.

Quando se faz amostragem em análise inferencial tem que ser considerado o grau de incerteza de que a amostra representa uma porção verdadeira da população que se pretende estimar. Desta maneira, uma quantidade de incerteza sempre segue a estimativa que se está fazendo. Esta incerteza pode ser calculada com técnicas de probabilidade.

Os procedimentos relacionados anteriormente são usados para descrever ou estimar valores de parâmetros de população baseados em amostragem aleatória. Porém há uma preocupação mais importante que procura entender a relação entre duas diferentes variáveis aleatórias, x e y , conhecida como relação bivariada.

É importante, neste ponto, a distinção entre variáveis experimentais e correlacionais. Na primeira nós podemos controlar o valor de uma variável e variar o valor da outra para entender o relacionamento entre ambas. Na segunda não há controle algum das variáveis sob estudo. Existem somente observações de como estas variáveis alteram se correlacionando no ambiente.

Esta distinção é importante porque em associações experimentais as relações causais podem ser feitas seguramente enquanto que em associações de correlação essas interpretações não tem o mesmo sentido.

Em condições formais, há uma relação hipotética descrita por

$$y = f(x) \quad (1)$$

onde o valor de y de uma observação em particular é uma função do valor de x daquela observação.

Existem padrões de associação entre variáveis. Eles podem ser positivamente relacionados, negativamente relacionados (ou inversamente). Eles podem estar positivamente em uma faixa e negativamente relacionados em outro faixa quando são descritos como não-monotônicos.

Quando não há alguma associação sistemática entre os valores de duas variáveis esta relação é chamada não relacionada, não-correlacionada, ortogonal ou independente.

A Reta de Regressão

A tarefa de determinar a presença ou ausência de uma relação entre duas variáveis aleatórias pode ser levada a cabo através de várias técnicas. Para usar todos os detalhes de observações emparelhadas, ao invés de uma mera inspeção visual, e ganhar perspicácia na relação bivariada, um gráfico ("scattergram") pode ser construído para os dados da amostra. Isto pode ser feito através da plotagem dos pares de observações da amostra em um gráfico bi-dimensional.

Quando duas variáveis aleatórias são mostradas em um gráfico algum tipo de relação aparece. Um modelo que indicasse uma relação linear entre x e y seria denominado um *modelo determinístico* - onde um determinado valor de x pode prever um valor para y .

Se esta fosse uma função perfeita, o valor exato de uma variável poderia ser estimado do valor da outra. Porém, outros fatores (ou variáveis) interagem com as duas variáveis observadas, o que induz ao aparecimento

de variância não explicada ou outros fenômenos aleatórios não facilmente modeláveis. Quando este erro do acaso é levado em conta temos:

$$y = f(x) + \varepsilon \text{ (erro aleatório)} \quad (2)$$

Assume-se que o componente determinístico representa o valor médio de y , com a notação $E(y)$. O componente aleatório permite que os valores de y tenham flutuação aleatória sobre a sua média.

O Coeficiente de Correlação r

A demonstração gráfica de duas variáveis aleatórias provê uma representação bi-dimensional pictorial dos dados. O coeficiente de correlação r é um índice numérico dessa relação. Ele sumariza a extensão da relação entre x e y .

O tipo mais comumente utilizado de coeficiente de correlação é o de Pearson – r , também chamado linear ou correlação de produto-momento. Ele provê uma medida quantitativa da força do relacionamento linear entre x e y . O valor da correlação (i.e., coeficiente de correlação) não depende das unidades de medidas específicas usadas, por exemplo, a correlação entre altura e peso será idêntica independente se polegadas e libras ou centímetros e quilogramas são usados como unidades de medida.

O coeficiente de correlação Pearson é calculado como segue:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y} \quad (3)$$

Em que $(x_i - \bar{x})$ é o desvio do valor da variável x da média daquela variável, $(y_i - \bar{y})$ é o desvio do valor da variável y da média daquela variável, S_x e S_y são os desvios padrões da amostra, e o somatório \sum é através dos $i = 1, 2, \dots$, pares de n . Um valor de r próximo ou igual a 0 implica

em pouca ou nenhuma relação linear entre y e x . Quanto mais próximo os valores são de 1, indicam uma força maior em uma relação linear positiva entre x e y . O mesmo acontece quando quanto mais próximo de -1, indicando a força de uma relação inversa.

É muito importante notar que uma correlação alta não implica em causalidade. A única informação obtida deste coeficiente é da existência de uma tendência linear entre as variáveis.

O Método dos Mínimos Quadrados

Com a informação provida pelos dados de amostra esta técnica permite ajustar um modelo linear aos dados. Alguns autores definem esta metodologia como um critério para escolher uma reta que melhor se ajuste aos dados plotados em um espaço cartesiano.

Basicamente o que o método faz é escolher uma reta particular, entre todas possíveis, isso resulta na menor soma dos desvios quadrados dos pontos dos dados da reta. Esta reta é chamada a *reta de regressão* ou *reta dos mínimos quadrados*, porque ela é determinada de uma tal maneira que a soma das distâncias quadradas de todos os dados da reta é a mais baixa possível.

Dado a fórmula do modelo probabilístico da reta:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (4)$$

em que y é a variável *dependente* a ser predita; x é a variável a ser usada como um previsor chamada de variável *independente*; β_0 (beta zero) é onde a variável y intercepta ou corta o eixo y ; β_1 (beta um) é a quantia que para cada unidade de aumento em x muda em y . Este é o mais simples dos modelos probabilísticos porque a parte determinística aparece plotada no gráfico como uma linha reta.

Quando a população inteira pode ser medida os parâmetros β_0 e β_1 podem ser calculados facilmente. Mas, na maioria das vezes, esta tarefa é

impossível. O que pode ser feito é, então, calcular estes valores de uma amostra de medidas. A solução matemática para achar estas estimativas é:

$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (5)$$

β_0 é uma função de, β_1 , \bar{x} , e \bar{y}

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (6)$$

a reta produzida por estes parâmetros (a reta dos mínimos quadrados) atravessará o ponto (\bar{x}, \bar{y}) .

Esta técnica escolhe a reta que minimiza a Soma dos Desvios Quadrados, $SSE = \sum (y - \hat{y})^2$, aonde \hat{y} é o valor previsto de y .

Algumas suposições têm de ser feitas quando se trabalha com um modelo deste tipo. A distribuição relativa de freqüência dos erros é considerada normal, com média zero e uma variância constante.

Para medir a variabilidade ao redor da reta dos mínimos quadrados a SSE é usada:

$$s^2 = \frac{SSE}{n - 2} \quad (7)$$

em que s^2 é a estimativa da variância do erro ou erro de estimativa. Calcula-se que a variação dos valores observados de y seja ao redor $2s$ do valor previsto \hat{y} .

O Coeficiente de Determinação

Quanto mais alto o grau de correlação entre as variáveis mais reduzidos são os erros de previsão. O coeficiente r^2 é uma medida da proporção de variância em uma variável responsável pela variação em outra. Pode ser usado para avaliar e reduzir os erros de previsão de y pela informação obtida de x . Este coeficiente pode ser determinado pela fórmula:

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} \quad \text{em que } 0 \leq r^2 \leq 1 \quad (8)$$

indica, como discutido antes, quanta variação da amostra de y pode ser atribuída quando se usa x para prever y . A distinção entre r e r^2 é que o primeiro mede a força da relação linear enquanto que o segundo indica como a reta dos mínimos quadrados se ajusta aos dados.

Regressão múltipla

Basicamente, o conceito de regressão múltipla é uma extensão do conceito de regressão simples. São usadas n variáveis independentes para prever a variável dependente. A forma geral da equação é:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (9)$$

em que y é a variável dependente, x_1, x_2, \dots, x_k , são as variáveis independentes.

Este modelo geralmente é denominado o modelo estatístico geral linear. Pode incluir algumas variáveis de ordem mais altas como por exemplo, $x_1^2, x_1 x_2$ ou x_3^3 .

A maioria dos cálculos em uma análise de regressão múltipla são muito elaborados e são executados, hoje em dia, em computadores. Os parâmetros, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$, devem ser determinados a partir de estimativas obtidas dos dados da amostra e devem ser ajustados ao modelo

para estes dados. A solução de mínimos quadrados é usada para determinar a melhor equação de regressão múltipla, um modelo calculado que tenta minimizar $SSE = \sum (y - \hat{y})^2$, em que \hat{y} é o valor predito de y .

A representação gráfica desta análise de regressão múltipla, em lugar de uma linha reta, é representada por planos em um espaço multi-dimensional, um conceito difícil de imaginar e impossível de descrever graficamente.

É importante distinguir que uma vez que, estes cálculos estão baseado em observações da amostra, eles são uma estimativa da verdadeira equação da população. Os procedimentos são feitos para calcular os parâmetros desconhecidos, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$, depois disso o modelo provido tem que ser examinado como um bom previsor ou não de y .

2.4 Ecologia da Paisagem

2.4.1 Introdução

O termo Ecologia da Paisagem foi desenvolvido inicialmente por um geógrafo alemão C.Troll (1950). Vários sinônimos são indicados na literatura para esse termo, tais como – Ecologia da Paisagem, Ecologia de Mosaicos da Terra, Ecologia Regional (Forman, 1995). Talvez o melhor modo para enfocar uma definição formal é buscar um entendimento do que a palavra paisagem significa.

Forman e Godron (1986) consideram que uma área heterogênea de terra composta de agrupamentos repetidos de ecossistemas pode ser definido como uma paisagem. Esta entidade é como um mosaico de tipos de cobertura terrestre (“patches” na tradução adaptada para – retalhos) formados pela inter-relação entre atividades humanas e os processos ecológicos. Estes mosaicos diferem em tamanho, forma e arranjo e eles representam, de alguma forma, o intercâmbio entre processos naturais e cultura humana.

Esse conceito de paisagem difere fundamentalmente do conceito tri-dimensional que estamos acostumados a imaginar. Nesse paradigma a paisagem é bi-dimensional, observada "por cima", apresentando os conjuntos de retalhos em uma superfície plana.

Quando a escala é a preocupação principal para uma definição, uma paisagem é uma superfície de terra espacialmente heterogênea comumente referenciada em escalas de hectares a muitos quilômetros (Gardner et al., 1987).

A combinação de fatores socioeconômicos e ecológicos determina padrões repetidos na superfície da terra. A interferência humana tem peso significativo moldando esse padrão da paisagem. Paisagens mais naturais ou menos influenciadas pelo homem mostram uma estrutura que significativamente difere daquelas que estão intensamente ocupadas pela cultura humana.

Podemos associar um certo grau de "naturalidade" relacionada às influências causadas ou não por atividades humanas. Os padrões da paisagem podem indicar muito das interações entre fatores socioeconômicos e ecológicos que acontecem na superfície da terra. Desenvolvimentos teóricos deram grande importância à relação entre padrão e processos e como esta informação pode ser extraída quando se varia a escala de trabalho (Gardner et al., 1991).

É importante distinguir o termo paisagem de uma sub-bacia ou bacia, de uma região, ou de um ecossistema (Forman, 1993). A sub-bacia ou bacia são delimitados por um rio ou riachos e seus tributários e os limites que delimitam uma bacia necessariamente não correspondem aos de uma paisagem. Uma região compreende a integração de várias paisagens. Um ecossistema tem um alcance muito grande de escala espacial e geralmente é uma área homogênea de terra.

Forman (1995) propõe alguns princípios gerais para a disciplina Ecologia da Paisagem. É importante destacar que estes princípios conceituais propostos incluem termos oriundos do campo de conhecimento da lógica difusa, onde o grau dos fenômenos são medidos pelas possibilidades de ocorrência, grande/pequeno; distinto/gradual; fino/grosso,

etc. Apesar disso, avaliações quantitativas determinísticas podem ser feitas baseadas em princípios aplicados quando descrições matemáticas dos princípios podem ser definidas.

2.4.2 Estrutura e Função da Paisagem

Uma paisagem é como um mosaico de retalhos da cobertura terrestre que são o resultado da interação de processos naturais e atividades humanas. Estes retalhos diferem em tamanho, forma e arranjo espacial (Lee et al. 1992). A paisagem, como um todo, é constituída de várias formas de elementos. O elemento mais abundante é a *matriz*. Este elemento dominante está mais conectado do que os outros, caracterizando assim a paisagem e influenciando a maneira como a mesma funciona.

Em uma escala temporal maior a influência do clima e dos processos geomorfológicos estabelecem um padrão na formação da estrutura da paisagem. Os regimes de perturbação, descritos a seguir, interagem com estes processos anteriores e definem a estrutura atual das paisagens.

Poderíamos classificar estes regimes de perturbações em naturais e antrópicos. O primeiro é representado, por exemplo, através de inundações, deslizamentos de lama, incêndios por raios, furacões, epidemias e o segundo através dos processos de preparo do solo, desflorestamento, aplicação de agrotóxicos, queimadas, mineração e outros.

A atividade humana previne a sucessão ecológica natural de acontecer interferindo constantemente na estrutura da paisagem. Indubitavelmente, a maioria da estrutura da paisagem na terra hoje foi influenciada pela atividade do ser humano.

A interação de todos estes processos resulta nos fatores causativos e explica em parte a origem do mosaico de retalhos. Porém nem todos esses processos têm uma origem causal, muitas vezes não há um padrão definido de formação e uma quantia considerável de aleatoriedade acontece.

A previsão da influência de agentes causais na configuração atual do mosaico de retalhos é uma tarefa difícil devido à influência de fatores externos à área do mosaico (Pickett e Thompson, 1978).

Vários processos acontecem em extensões de escalas espacial e temporal muito diferentes e extensas, o que faz com que estudos de previsão sejam difíceis de modelar. A dinâmica da paisagem é, talvez, muito mais influenciada por atividades humanas do que por perturbações naturais.

Retalhos

A maioria dos retalhos pode ser classificada considerando-se o modo como eles foram formados. Porém, como foi comentado anteriormente, uma quantia considerável de aleatoriedade acontece na natureza e, em função disso, nem todos os relacionamentos causais podem ser explicados.

Vários aspectos podem ser reconhecidos quando uma tentativa é feita para classificar os retalhos. Tamanho ou área são os aspectos identificáveis à primeira vista. A maioria das decisões de alocação de terra são feitas considerando o melhor tamanho de retalho para uma certa atividade.

Quando uma paisagem é classificada de acordo com sua configuração de retalhos (origem, tamanho, forma) isto não significa que um padrão espacial possa ser definido. Em outras palavras, técnicas estatísticas são necessárias para entender a distribuição espacial dos retalhos (aleatória, regular, agregada).

A estrutura da paisagem pode ser descrita por índices (Fung, 1987), tais como, a quantidade de extremidades do retalho, a densidade de retalhos, e a distância entre retalhos.

Várias medidas de estrutura da paisagem foram propostas, Baker e Cai (1992) propõem algumas medidas para analisar quantitativamente a estrutura da paisagem para escalas espaciais muito pequenas. Elas são:

(1) idade média do retalho - idade média de todos os retalhos na paisagem;

(2) tamanho médio do retalho - tamanho médio de todos os retalhos na paisagem;

(3) distância média - distância média, medida do centro de um retalho ao centro de cada retalho que compartilha um limite com aquele retalho;

(4) número médio de pixels (“picture element”) da extremidade - é uma contagem do número de pixels que encontram-se no limite entre os retalhos na paisagem;

(5) densidade dos retalhos - contagem dos retalhos na área;

(6) diversidade de idade dos retalhos - é uma medida combinada da igualdade e riqueza de idade dos retalhos;

(7) segundo momento angular - medida da textura da escala da paisagem;

(8) fração de crescimento velho - é a fração da área total de uma paisagem que é coberta por retalhos cujas idades são maiores que 200 yr.

2.4.3 Modificação da Paisagem

Forman e Godron (1986) mostraram que alguns fenômenos causam modificações nas paisagens em menos de um dia, por exemplo, incêndios de florestas, inundação fluvial, etc. Porém outras perturbações acontecem em níveis maiores na escala temporal e espacial. Estas causas podem ser a expansão das cidades, desertificação, irrigação, desmatamento, reflorestamento e outros.

Eles argumentam que é possível determinar quais parâmetros fazem a paisagem variar durante o tempo, apesar das diferenças de escala. Acabam propondo, independente da escala temporal, uma curva de variação que descreve a relação de modificação da paisagem ao longo do tempo.

A curva de variação tem três parâmetros independentes:

1. uma tendência geral (aumentando, diminuindo);
2. amplitude de oscilação;
3. ritmo de oscilação.

Eles listaram características de mudanças importantes que podem ser estudadas em uma paisagem tais como formas dos retalhos, porosidade da matriz, desenvolvimento das cadeias ecológicas, produtividade, níveis de nitrogênio, diversidade biótica, espécies de pássaros raras, níveis de sucessão vegetal, fluxos de matéria, biomassa total e assim sucessivamente.

A curva de variação indica se uma paisagem é estável ou não. É estável quando uma variação a longo prazo é representada por uma linha horizontal. Na verdade isto pode ser considerado como uma meta-estabilidade, onde o sistema oscila ao redor de uma posição neutra.

Destes conceitos, *instabilidade* pode ser definida como qualquer mudança que diverge desta tendência geral em uma linha horizontal. A instabilidade pode ser temporária ou persistente.

Os autores também indicam que uma terminologia caótica é apresentada na literatura para definir a estabilidade de sistemas ecológicos. Não obstante eles propõem *persistência*, como uma medida de estabilidade, representando uma característica presente de uma paisagem sob um certo período de tempo.

2.4.3.1 Medidas de Modificação da Paisagem

A estrutura de paisagem atual, como discutido anteriormente, foi formada pela interação do clima, geomorfologia, perturbações naturais e artificiais. Esta estrutura não está completamente em um estado de estabilidade. A dinâmica em uma paisagem pode ser interpretada basicamente como o encontro de processos de sucessão ecológica naturais contra a luta do homem para manter essa paisagem culturalmente habitável.

A questão central é identificar como uma paisagem muda, como isso acontece e, se há um padrão de mudança ou não.

Um modo prático para distinguir este mosaico de trocas é a análise dos retalhos como fases diferentes de um fluxo de mudanças. Os tipos de elementos da paisagem mudam através do tempo indicando uma sequência e uma proporção dentro desta mudança.

Para entender a dependência entre as muitas variáveis ecológicas e sociais que formam um padrão da paisagem, foram desenvolvidos vários métodos. A análise quantitativa destes processos requer a pesquisa e desenvolvimento de medidas quantitativas do padrão da paisagem.

A modificação da paisagem é uma questão complexa devido à diversidade dos processos na paisagem. Muitos tipos e formatos de dados

podem ser usados para medir a modificação do padrão da paisagem. Vários conceitos têm que ser considerados para analisar essa modificação da paisagem. Uma variedade de fontes de dados e tecnologias são usadas para avaliar as modificações, e, para relacionar toda essa informação, uma metodologia de classificação tem que ser aplicada.

Uma variedade extensa de índices foi desenvolvida para entender as modificações que são descritas nos padrões da paisagem.

2.4.3.2 Métodos Quantitativos

Turner e Gardner (1994) indicaram a necessidade de métodos quantitativos para unir padrões espaciais e processos ecológicos em grandes escalas espaciais e temporais.

Estes métodos quantitativos podem analisar padrões, investigar processos ecológicos e desenvolver modelos de modificação da paisagem.

Os processos e consequentemente os problemas ecológicos alcançam uma extensão muito grande por causa de suas escalas espaciais e temporais. As previsões têm que ser feitas inferindo-se de diferentes escalas tanto temporais como espaciais.

Estudos mais recentes indicam que informação pode ser extrapolada através das escalas. Uma paisagem tem limites críticos onde modificações qualitativas acontecem em seus processos ecológicos.

Estes estudos também indicam que, ao nível da paisagem, algumas variáveis – *variáveis-chave* – são mais importantes para predizer os padrões da paisagem ou processos do ecossistema. Isto poderia simplificar a previsão da dinâmica de modificação da paisagem. Apesar destas considerações, previsões ou inferências sobre os padrões da paisagem ou processos ecológicos devem ser feitas sempre considerando a escala de estudo.

São descritos vários métodos quantitativos na literatura para análise e interpretação de padrões da paisagem. Aplicação de estatística espacial; a capacidade do Sensoriamento Remoto e dos GIS (Sistemas de Informações

Geográficas); a aplicação de medidas de textura; a combinação de GIS e inteligência artificial; os fractais, para análise da complexidade dos padrões da paisagem.

2.4.3.3 Modelos de Modificação da Paisagem

Quando se estuda modificações a nível da paisagem isto necessariamente conduz a previsão de fenómenos. A previsão de fenómenos por sua vez conduz ao desenvolvimento de modelos. Estes modelos são importantes porque experimentos não podem ser executados devido a magnitude das escalas reconhecidas nos estudos da paisagem.

Baker (1989) revisando modelos de modificação da paisagem mostrou as suposições, limitações e possíveis aplicações dos mesmos. Ele classificou estes modelos pelo critério do nível de agregação aonde o elemento da paisagem é o valor a ser agregado no modelo. Ele distinguiu três tipos de modelos baseado neste critério:

1. Paisagem Inteira → o valor de uma variável em alguma área da paisagem é modelado; 2. Modelos da Paisagem Distributivos → a distribuição de valores de uma variável em alguma área da paisagem é modelada; 3. Modelos da Paisagem Espaciais → a circunstância de sub-áreas individuais na paisagem e sua configuração é modelada usando como sub-modelos um conjunto de modelos descritos anteriormente.

O autor indica que modelos espaciais usam a localização e a configuração de elementos da paisagem quando projetam modificações. As regras dos modelos espaciais são: a escolha entre os formatos raster ou vetor de representação de dados, que são as variáveis no modelo; o tamanho ou resolução do pixel; e, finalmente, o algoritmo de modificação da paisagem.

Existem basicamente dois tipos de modelos espaciais. O de mosaico aonde a paisagem é dividida em células (em pixels ou polígonos) e a coleção destas células formam o mosaico. Ele pode representar uma ou muitas variáveis com um ou muitos planos de células. O segundo tipo de modelo espacial é o modelo de elemento. Neste, o tipo de elemento da paisagem é

modelado para modificar. Estes modelos têm a mesma representação do prévio, uma base de dados de malha de células ou vetorial da localização dos tipos de elementos mapeados.

Baker (1989) considerou que estes tipos de modelos não são modelos da paisagem porque eles se preocupam mais com organismos individuais (árvores de floresta; organismos marinhos). Porém algumas analogias podem ser feitas nestes modelos tornando-os utilizáveis para o estudo da modificação da paisagem. Elementos da paisagem individuais poderiam ser considerados como organismos individuais examinando o crescimento dos mesmos, a sua composição, idade ou características físicas.

Apesar de considerar que não há nenhum modelo de modificação da paisagem perfeito, várias tendências e esforços de pesquisa podem modificar este fato em um curto período de tempo.

À medida que dados espaciais estão ficando mais e mais disponíveis e a capacidade computacional vem aumentando rapidamente, modelos espaciais deverão tornar-se mais convenientes como uma ferramenta científica e de gestão ambiental.

É importante salientar que uma grande limitação ocorre quando se faz estudos de modificação da paisagem: a obtenção de dados. Fotografias aéreas estão disponíveis somente a partir dos anos trinta e quarenta. Dados de Satélite desde a década de setenta.

Entender como e por que as paisagens modificam é o desafio mais importante para o desenvolvimento de novos tipos de modelos. A importância da manutenção de paisagens inalteradas é enorme para o desenvolvimento de pesquisas que expliquem os processos naturais. Os processos da paisagem devem ser estudados empiricamente para que se possa usar os dados resultantes destas pesquisas nos modelos de modificação da paisagem.

2.5 Sistemas de Informações Geográficas (GIS)

2.5.1 Histórico

A representação pictórica da superfície da terra tem sido desenvolvida e usada através da história da humanidade desde os tempos antigos até os tempos modernos (Burrough, 1986). A ciência da cartografia evoluiu em consonância com a da navegação e com propósitos militares. Levantamentos terrestres para reconhecimento de terreno e planejamento agrícola também contribuíram significativamente para este desenvolvimento.

Mais tarde, à medida que várias disciplinas científicas foram progredindo, como a geologia, a geomorfologia, a ecologia e outras e, a necessidade pela informação mais precisa intensificou-se, os esforços para o mapeamento dos lugares de interesse tornaram-se mais específicos.

Vários campos do conhecimento contribuíram provendo ferramentas novas, tanto intelectual como metodologicamente, para a análise de dados espaciais (dados com uma localização geográfica). Disciplinas como estatística, teoria dos números e matemática avançada contribuíram significativamente para esta evolução (Star & Estes, 1990). A revolução industrial, com demanda forte por matérias-primas e trabalho, causou uma forte exigência para uma "infra-estrutura nova, extensa, tanto social como industrial". Mapas que descrevem população, estradas, indústrias locais, locais urbanos são agora de interesse público e do setor privado. Enfoques quantitativos são desenvolvidos para resolver "problemas-baseados-em-mapas", ou seja, os mapas deixaram de ser uma ferramenta apenas consultiva para fazer parte no processo de tomada de decisão.

Com o advento do computador e a transição da capacidade de manipulação da informação analógica para a digital, a capacidade para armazenar e analisar dados aumentou enormemente. Três fatores principais influenciaram o aparecimento dos Sistemas de Informações Geográficas (GIS): 1. técnicas cartográficas mais precisas; 2. sistemas de computadores digitais e 3. a revolução quantitativa na análise espacial (Star & Estes, 1990).

Goodchild, (1993) informa que talvez o primeiro uso sistemático de computadores na análise de mapas foram utilizados no Canadá, para inventariar o país. A maioria da análise foi conduzida para a medição de áreas. A análise por computador provou ser a única maneira de realizar uma tarefa enorme e incômoda com uma relação custo-benefício efetiva.

Outra capacidade importante dos GIS é a habilidade para sobrepor dois ou mais mapas para análise. Naquela época uma quantia significativa de mapas de recursos unidisciplinares estava disponível e esforços foram feitos para a integração destas informações. Neste começo, sobrepunham-se cópias transparentes em uma mesa iluminada e a informação extraída podia por exemplo representar as semelhanças entre os atributos estudados. Foram desenvolvidos vários programas com capacidade de sobrepor vários mapas e extrair informações do processo.

2.5.2 Elementos Essenciais

Não há consenso algum na literatura atual sobre uma definição completa para o que seja um Sistema de Informações Geográficas. Vários exemplos são apresentados abaixo em ordem cronológica:

" ...um conjunto poderoso de ferramentas para colecionar, armazenar, recuperar, transformar e exibir dados espaciais do mundo real para um conjunto particular de propósitos " (Burrough, 1986).

" ...qualquer conjunto de procedimentos manuais ou baseados em computador usados para armazenar e manipular dados geograficamente referenciados."

" ...uma coleção organizada de hardware e software de computador, dados geográficos e recursos humanos projetados para capturar eficazmente, armazenar, atualizar, manipular, analisar, e exibir todas as formas de informação geograficamente referenciada" (ESRI, 1995).

" ...um sistema para preparar, apresentar, e interpretar fatos que pertencem à superfície de terra " (Tomlin, 1990).

" ... um sistema assistido por computador para a aquisição, armazenamento, análise e exibição de dados geográficos " (Eastman, 1995).

Operações de dados como coleta, armazenamento, manipulação, análise e exibição estão presentes na maioria das definições. Baseada no computador e geograficamente referenciada são características consideradas importantes para uma definição formal.

Além das definições formais, vários autores concordam com o fato que os GIS estão ficando cada vez mais complexos, e para ser considerado um GIS, um sistema deve ter algumas funções essenciais mínimas. Porém, também não existe consenso algum entre estes autores sobre o conjunto mínimo necessário de capacidades para definir um GIS.

Alguma sistematização (Star & Estes, 1990) tem sido feita e alguns elementos funcionais essenciais podem ser elencados. *Aquisição de dados* -- normalmente envolvendo observações de campo, interpretação de fotografia aérea, dados de censo. *Pré-processamento* -- representado pela manipulação de dados para torná-los utilizáveis para um GIS. Os processos principais são formatação e geocodificação de dados. *Administração de Dados* -- representando a criação e acesso ao banco de dados do sistema. *Manipulação e Análise* -- são executadas operações analíticas e informação nova é derivada interagindo com o banco de dados. Os autores indicam que o termo *geoprocessamento* está normalmente associado com este elemento funcional. *Geração de produto* - produtos finais tais como relatórios estatísticos, mapas, gráficos.

Burrough (1986) define os elementos essenciais de um GIS com três componentes - hardware de computador, módulos de software e contexto organizacional. Mais tarde ele define os "cinco módulos" técnicos básicos como: 1. Introdução e verificação de dados; 2. armazenamento de dados e administração de banco de dados; 3. produção e apresentação de dados; 4. transformação de dados e 5. interação com o usuário.

Eastman (1995) argumenta que para ser considerado um "verdadeiro" GIS um sistema tem que conter pelo menos: *Banco de dados espaciais e de atributos* -- informação em configuração digital sobre mapas e atributos correlatos; *Sistema de exibição cartográfico* -- saída de resultados adquiridos através de imagens ("soft copies") ou gráficos, relatórios ou mapas ("hard copies"); *Sistema de administração de banco de dados (DBMS)* -- é um conceito mais amplo que os sistemas de administração de banco de dados tradicionais porque são incluídas capacidades para administrar dados espaciais e atributos; *Sistema de análise geográfica* -- habilidade para

analisar dados considerando suas características espaciais; *Sistema de processamento de imagens* -- mais recentemente alguns programas de computador incluem módulos que processam imagens e executam algumas das tarefas relacionadas à imagens de sensoriamento remoto; *Sistema de análise estatística* -- rotinas tradicionais e específicas para a descrição estatística de dados espaciais; e *Sistema de apoio a decisão* -- freqüentemente reivindicados como um dos maiores benefícios dos GIS, este módulo não está presente na maioria dos programas.

É interessante indicar que nenhum sistema sozinho contém todas as capacidades analíticas que se possa precisar. No mesmo raciocínio que livro algum possa cobrir todos os assuntos relacionados aos GIS.

Talvez o ponto central nesta discussão é que o GIS se tornou um campo tão imenso em tecnologia, conhecimento e integração que tentativas para descrever o sistema como um todo se tornaram incompletas e infrutíferas (Nyerges, 1993).

2.5.3 Estrutura dos Dados

O fator principal sobre os dados manipulados por um GIS é que eles têm uma posição georeferenciada, conexões topológicas e suas características associadas são normalmente armazenadas em uma forma tabular (Burrough, 1986).

Pelo fato de um banco de dados de computador (assumindo que os GIS sejam computadorizados) ser um repositório finito e discreto de dados, Goodchild (1993) argumenta que é necessário, para modelar um fenômeno do mundo real, condensar, abstrair e generalizar a informação. Fazendo assim, a superfície da terra pode ser representada por objetos ou entidades - na maioria das vezes descritas como conceitos topológicos ou os três tipos primitivos de entidades geométricas - pontos, linhas, e polígonos (ou planos, ou áreas).

É importante estabelecer que dados geográficos ou espaciais não significam somente dados cartográficos e científicos mas também dados empresariais, registros de terra, fotografias, bancos de dados de clientes,

guias de viagem, registros imobiliários, documentos legais e assim por diante. Em torno de 80% dos dados usados em negócios e pelo governo estão relacionados (ou podem ser relacionados) à características geográficas (códigos postais, endereços, localizações de fábrica, áreas de censo, cidades, estados e países).

Estes dados espaciais podem ser representados em um ambiente de computador basicamente por duas formas de armazenamento, apesar de alguma controvérsia de definição na literatura. A maioria dos pacotes de software em uso hoje utilizam os formato *vetor* e/ou *raster*.

Os dados *raster* registram informações espaciais em uma malha regular ou matriz composta por linhas e colunas. Cada célula (também denominadas de pixels ou elementos de quadro) nesta malha contém um atributo que pode representar alguma característica espacial específica. Este tipo de dado, considerado como o de estrutura mais simples, é manipulado mais facilmente pelos computadores porque as sequências de filas e colunas, podem ser armazenadas, manipuladas e exibidas (Burrough, 1986).

Os dados *vector* armazenam dados espaciais em coordenadas x,y em um sistema planar de coordenadas. São registradas características pontuais como coordenadas x,y distintas. Linhas e polígonos são registrados como uma sucessão ordenada de coordenadas de x,y . A **tabela 1** apresenta uma comparação entre esses dois tipos de armazenamento.

2.5.4 Capacidades analíticas

O que distingue um GIS de um CAD (computer aided design) e outros pacotes de software semelhantes são que o primeiro pode transformar os dados espaciais originais para a execução de tarefas específicas e questões particulares (Burrough, 1986).

Independentemente da representação da estrutura de dados a maioria dos pacotes de GIS hoje contém módulos de análise com vários algoritmos para execução de uma variedade de funções dentro da base de dados disponível.

Uma lista de 371 programas de GIS é apresentada no periódico GIS World, Inc. (1991) com um vasto alcance de capacidades analíticas e desenvolvido para várias plataformas.

Provavelmente dentre a maioria dos múltiplos usos dos GIS a cartografia temática representa a maior utilização através da composição de mapeamento multicoloridos. Goodchild (1993) estabelece uma distinção clara entre a atividade de GIS e a de cartografia digital. Berry (1993) apresenta uma lista indicando o que é necessário para um GIS ir além do simples mapeamento, em outras palavras, a modelagem de dados espaciais.

Estas capacidades analíticas seriam: *operadores matemáticos* – matemática básica e avançada, e linguagem de macro comandos; *estatística espacial* – descritiva, comparativa, preditiva; *medição de distância* – distância simples, tamponamento, estreiteza, proximidade simples, proximidade efetiva (caracterização de movimento); *caracterização de vizinhança* – configuração de superfície, resumo de janela móvel, interpolação, delineação de ponto de vista, densidade de exposição; *otimização de caminhos* – caminhos simples, geometria de fractais, integridade espacial, proximidade, distância de interface.

A **figura 1** apresenta um fluxograma de um processo de trabalho típico de um GIS. O autor indica que alguma sobreposição sempre acontece entre as diferentes fases.

Quando se averigua os usos dos GIS através de um grau crescente de complexidade (**figura 2**) pode-se iniciar com um simples inventário e administração dos dados para sofisticadas análises e modelagem de dados espaciais (Goodchild, 1993).

2.5.5 Modelagem Ambiental e GIS

Através da simulação torna-se possível o entendimento do processo de projetar um modelo e a condução de experiências com o propósito de entender o comportamento de um determinado sistema. Um modelo é uma representação de um grupo de objetos ou idéias diferentes da realidade que está sendo modelada.

| <i>RASTER</i> | <i>VECTOR</i> |
|--|---|
| vantagens | vantagens |
| 1. estrutura dos dados simples | 1. boa representação da estrutura fenomenológica dos dados |
| 2. operações de sobreposição e combinação de dados mapeados através de sensoriamento remoto é mais simples | 2. estrutura de dados compacta |
| 3. vários tipos de análises espaciais são facilmente realizáveis e a variabilidade espacial é eficientemente representada | 3. suporta dados gráficos mais eficientemente |
| 4. simulações são mais eficientemente conduzidas pelo fato de cada unidade espacial ter a mesma forma e tamanho | 4. melhor eficiência aonde informação topológica se faz necessária |
| 5. a tecnologia é barata e grande esforço tem sido despendido no seu desenvolvimento | 5. recuperação, atualização e generalização de dados gráficos e atributos é possível |
| 6. o mesmo conjunto de células da malha pode ser usado para diversas variáveis | 6. extensamente utilizado para descrever regiões administrativas |
| 7. mais simples para programar | |
| desvantagens | desvantagens |
| 1. uso dispendioso da capacidade de armazenamento para dados espaciais esparsos | 1. estrutura de dados complexas |
| 2. erro na estimativa de forma e perímetro | 2. simulação difícil de ser executada pelo fato de cada unidade ter conformação topológica diferenciada |
| 3. difícil estabelecimento de correlações quando se trabalha com redes ("networking") | 3. sobreposição de vários polígonos torna-se lenta e custosa |
| 4. transformações das projeções são extremamente custosas | 4. plotagem podem ser muito caras |
| 5. o uso de células grandes para a redução do volume de armazenamento de dados pode conduzir à perda de estruturas fenomenológicas reconhecíveis e perda de informação | 5. a tecnologia é cara |
| 6. mapas nesse formato são geralmente menos precisos e de bom acabamento | 6. variabilidade espacial não está implicitamente representada |

Tabela 1: Comparação entre dados Raster e Vector

Fonte: Adaptado de vários autores

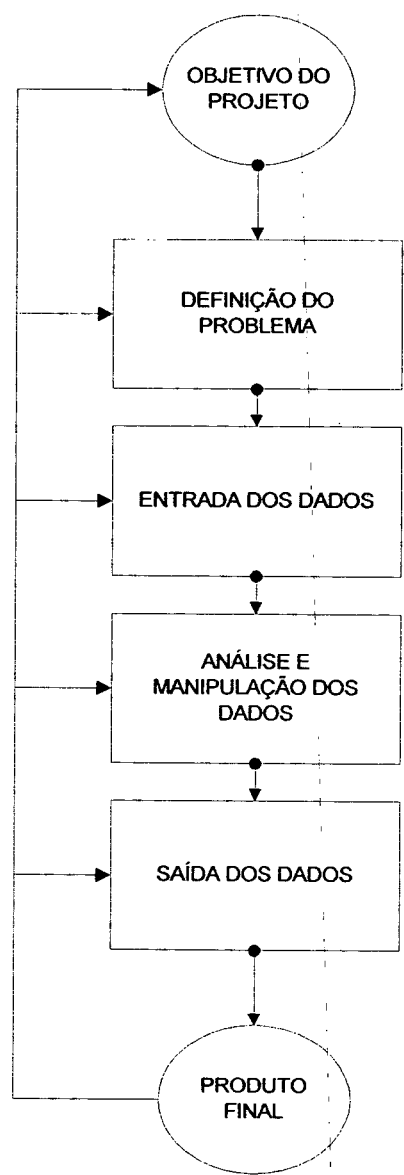


Figura 1: Perspectiva do procedimento nos GIS
Fonte: Nyerges, 1993

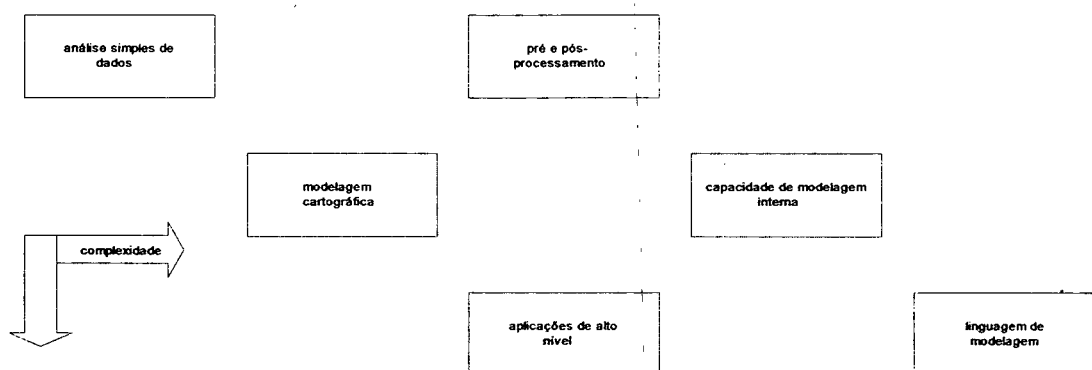


Figura 2: Grau crescente de complexidade para os GIS

Fonte: Adaptado de vários autores

Como discutido na seção anterior o processo de discretização do mundo real conduzido por um GIS condensa, abstrai e generaliza os dados espaciais. Dados espaciais são de natureza complexa. Dependendo da escala de aproximação (temporal ou espacial) podem ser extraídos e interpretados vários tipos de informações destes dados.

Para entender melhor o mosaico de diferentes fenômenos que acontecem na vida real é necessário projetar um sistema que simule esta realidade. Os GIS contêm muitas características que podem realizar grande parte destas tarefas. Uma tendência geral de complexidade foi mostrada (**figura 2**) e o modo que os GIS executam esta simulação de fenômeno do mundo real. De tarefas triviais de análise de dados para o que poderia ser definido como "modelagem verdadeira".

Vários enfoques integrando os GIS e a Modelagem Ambiental estão extensamente discutidos em Goodchild (1993).

Um conceito básico em GIS é o de localização (distribuição e relacionamentos) e objetos espaciais são os componentes básicos enquanto que na Modelagem Ambiental os conceitos básicos são de estado, de interação e dinâmica (números, massa, energia) e os componentes básicos são biológicos, químicos e físicos (espécies, ar, água, sedimento) (Fedra, 1993).

Uma vez que ambas unidades básicas nos GIS e na Modelagem Ambiental têm uma distribuição e localização espacial um paradigma que funda os dois campos é uma ferramenta muito promissora para avaliar problemas ambientais.

Ambos os campos têm os próprios esforços de pesquisa para o desenvolvimento de ferramentas de avaliação ambiental. Aplicações Ambientais em GIS se preocupam com tecnologia de informação; sensoriamento remoto como uma fonte de dados; integração com sistemas especialistas; sistemas de apoio à decisão; modelagem de água superficial, e assim por diante.

A Modelagem Ambiental é um campo do conhecimento mais antigo com vários sub-campos e uma quantia enorme de pesquisa e desenvolvimento empreendidas no último século. Várias sub-divisões do campo existem: sistemas atmosféricos; sistemas hidrológicos; superfície da terra e processos de sub-superfície; modelos biológicos e ecológicos; análise de risco e perigos entre outros (Fedra, 1993).

Pode haver graus diferentes de integração. A **figura 3** representa como a integração pode acontecer em níveis diferentes.

No primeiro há somente uma troca de arquivos (informação). Neste nível o que basicamente acontece é uma modificação de formatos de arquivo e rotinas de entradas/saídas de dados. O GIS é usado para armazenar e gerenciar os dados de entrada do modelo e, mais tarde, mostrar a integração dos resultados. Mais integração significa compartilhamento de memória e arquivos junto com uma interface de usuários comuns. Este grau de integração normalmente é alcançado com uma linguagem de aplicação de alto-nível.

O conceito de arquitetura aberta dos GIS conduzirá esta tendência de integração permitindo que tipos diferentes de programas funcionem juntos como um único sistema.

Mais recentemente alguns pacotes de software incluíram várias função analíticas em seus módulos o que tem tornado possível várias

aproximações de modelagem ambiental. O modelo, ou parte do mesmo, torna-se uma função analítica do GIS.

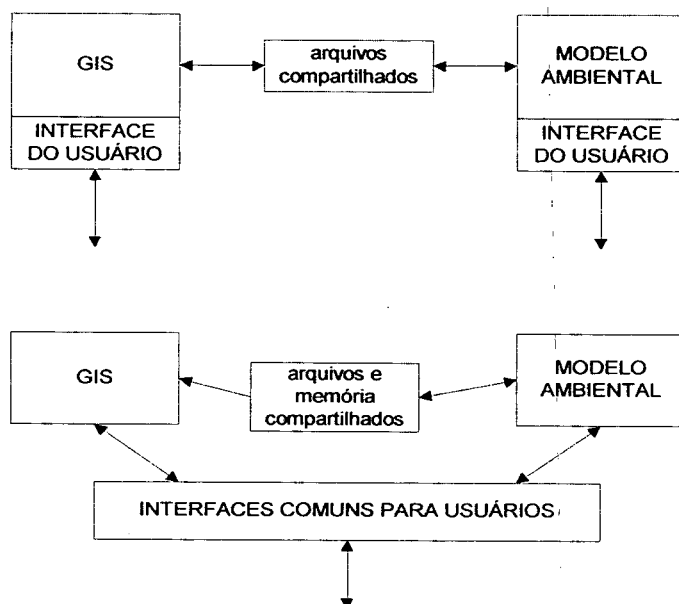


Figura 3: Integração entre os GIS e os Modelos Ambientais

Fonte: Fedra, 1993

Fedra (1993) imagina um sistema de GIS no futuro tão fácil de usar pelo fato de "...não ser estruturado em termos de variáveis de estado e parâmetros, ou objetos espaciais e atributos, mas em termos de problemas, intenções e objetivos, restrições e regulamentos, opções e alternativas de decisão, fatos e suposições, preferências e percepções. "

2.6 Sensoriamento Remoto (SR)

2.6.1 Definição de Sensoriamento Remoto

Um processo que envolva a coleta de dados de um objeto, uma área, ou um fenômeno sem contato físico com estas respectivas fontes, juntamente com a interpretação destes dados para extrair informação, é definido como

Sensoriamento Remoto (SR) (Lillesand, 1994). Qualquer sistema que derive informação e não esteja em contato com os objetos ou fenômenos de interesse (Star et al., 1990).

Mais especificamente essa definição tem de estar associada com a medição de interações entre os materiais da superfície da terra e a energia eletromagnética (Eastman, 1995).

Os produtos de qualquer representação pictórica em Sensoriamento Remoto são a fotografia e a imagem eletrônica. Frequentemente o termo imagem é usado para representar os dados de qualquer cena.

O propósito da obtenção de dados através do Sensoriamento Remoto é para a identificação e avaliação dos materiais na superfície e suas propriedades espaciais (Richards, 1993).

A maioria dos dispositivos de SR utiliza energia eletromagnética. Os sensores podem ser passivos ou ativos em relação à medição dessa energia. Os primeiros, mais comuns, medem os níveis de energia no ambiente onde geralmente o sol é a maior fonte. Resumidamente, os sensores passivos são aqueles que não emitem a energia que está sendo detectada (Eastman, 1995).

Em relação aos sensores ativos eles provêm a própria fonte de energia e o exemplo mais familiar é o "flash" utilizado em máquinas fotográficas. Os radares emitem energia na região das microondas no espectro eletromagnético

O fato do espectro eletromagnético ser muito extenso faz com que somente uma faixa de comprimentos de onda sejam utilizados pelo Sensoriamento Remoto. Na verdade é a diferente interação entre os comprimentos de onda e os materiais na superfície da terra que tornam o SR efetivo ou não.

Por causa das grandes distâncias da maioria dos sensores de SR vários processos afetam o caminho da energia eletromagnética. Fatores tais como, condições atmosféricas (névoa); problemas mecânicos, ópticos e elétricos no sensor; transmissão de dados; contribuem para alterar a resolução radiométrica (valores de brilho dos pixels).

A atmosfera causa significativa interferência nas ondas eletromagnéticas absorvendo e espalhando os comprimentos de onda mais curtos (Eastman, 1995). Na verdade a luz dita visível é a faixa de onda que ultrapassa a atmosfera com menos interferência.

O número e localização das medidas espectrais têm grande valor para investigação e manipulação porque cada material tem seu próprio padrão de resposta espectral (também denominado assinatura espectral). Os olhos humanos não podem perceber todo os comprimentos de onda do espectro luminoso, este fato faz a análise multispectral por SR muito útil porque informação aparentemente não disponível pode ser medida pelo sensor.

Considerando o fato que nem todos os comprimentos de onda do espectro eletromagnético são utilizáveis para os objetivos do SR, principalmente por causa da interferência da atmosfera e limitações tecnológicas, a maioria das aplicações são conduzidas na faixa do verde, vermelho e infra-vermelho-próximo (IR) (entre aproximadamente 0.4 e 12 μm) para sobrepor tais inconveniências.

Richards (1993) argumenta que o significado dessas faixas de utilização diferentes dizem respeito às interações distintas entre a radiação eletromagnética e as características dos materiais que estão sendo analisados.

Na faixa do visível e do infra-vermelho a energia medida depende de características dos materiais sendo medidos tais como, pigmentação, estrutura celular e grau de umidade da vegetação. Na faixa termal do infra-vermelho (no final) o calor emitido pelos objetos é a característica mais importante a ser medida. Na faixa das microondas, onde se utiliza sensores ativos a rugosidade da superfície e suas propriedades elétricas são fatores determinantes na magnitude do sinal refletido.

Essas diferenças na potencialidade de fornecer informação que cada comprimento de onda tem levou ao desenvolvimento de sensores especializados em determinadas faixas de comprimento de onda, complementando a informação ou sendo complementado por outros sensores.

Essas diferenças levaram ao desenvolvimento de sistemas de SR que podem tomar amostras dentro do espectro de comprimento de ondas.

2.6.2 Processamento Digital no Sensoriamento Remoto

Os dados de imagem da superfície da terra são acessíveis em formato digital compostos de elementos discretos, ou pixels, e com atributos de níveis de brilho (quantificação radiométrica). Estes dados de imagem tem uma resolução espacial (Richards, 1993).

A resolução espacial, do ponto de vista de um sensor do sistema, é a habilidade de tornar uma imagem nitidamente definida. Três níveis de resolução se sobrepõem de alguma maneira. O nível de descoberta onde um objeto de interesse é primeiramente distinguido; em seguida o nível de reconhecimento onde este objeto é categorizado; e finalmente o nível de identificação onde mais informação é adicionada. A **tabela 2** apresenta exemplos de resolução espacial (solo) de alguns sensores de SR.

A grande vantagem da obtenção dos dados no formato digital é o seu processamento imediato por computador. Operando digitalmente o computador pode distinguir 256 níveis de cinza enquanto que o olho humano somente pode perceber entre oito a dezesseis níveis.

Além da capacidade de distinguir melhor os níveis de cinza o processamento digital pode extrair uma quantidade enorme de informação espectral presente em uma imagem de satélite que um ser humano, por mais especialista que fosse na matéria, poderia fazer. Como exemplo, uma cena do satélite Landsat TM contém mais de 200 megabytes de dados.

Richards (1993) indica duas formas de se extrair informação das imagens provenientes do SR. O uso do computador para analisar digitalmente o pixel que ele denomina *análise quantitativa*. A outra forma é um analista humano fazendo uma inspeção visual para extrair informação e tem a denominação de *fotointerpretação*. Ambos os enfoques tem as suas próprias regras e são complementares. De uma maneira resumida pode-se dizer que a análise digital tem mais precisão quantitativa enquanto que a

| Dispositivo | Resolução no solo |
|---|---|
| NOOA (advanced very high resolution radiometer) | 1.1 x 1.1 Km |
| Landsat (earth resource satellite) | 30 x 30 m (bandas 1 - 5 e 7) 120 x 120 m (banda 6 - termal) |
| SPOT (Système Pour L'Observation de la Terre) | 20 x 20 m (banda 1 a 3) 10 x 10 m (banda pancromática) |

Tabela 2: Exemplos de Resolução Espacial

Fonte: modificada de diversos autores

análise humana tem mais precisão qualitativa.

O objetivo tanto da análise digital como da humana são basicamente: detectar, identificar (ou classificar), e medir as características das feições (Star et al., 1990). O processo digital baseia-se principalmente em tons de cor diferentemente do processo visual humano que incorpora elementos da imagem.

A característica mais marcante da análise digital é a sua consistência e possibilidade de repetição constante. Desta maneira massivas quantidades de dados podem ser explorados e analisados. O fato do processo de analisar uma imagem digital ser altamente subjetivo torna ainda mais falho a fotointerpretação pois é muito mais improvável que um analista possa repetir continuamente sua interpretação da mesma maneira que uma máquina poderia imparcialmente fazer.

Star (1990) enfatiza que nenhuma é definitivamente superior à outra. A capacidade de discernimento do ser humano torna possível a incorporação de fatores na interpretação da imagem que o processo assistido por máquina não pode desenvolver.

Convém ressaltar dentro do tema processamento digital a operação de correção geométrica onde os dados da imagem são registrados em um sistema de coordenadas diferentes para a mesma cena de interesse (usualmente as imagens de satélite não estão registradas em sistema algum)

O processo usa equações polinomiais para estabelecer uma “rubber sheet transformation” como se a imagem fosse colocada em uma borracha e modificada até que correspondesse àquela em que está sendo referenciada. Um conjunto de equações polinomiais é desenvolvido para descrever o mapeamento espacial dos dados antigos para os novos dados do sistema referencial.

2.6.3 Sensoriamento Remoto e Sistemas de Informações Geográficas

A maioria dos dados organizados nos GIS (exceção para aqueles que possuem também capacidade de manipulação raster) constituem-se de conjuntos de categorias discretas (tipos de vegetação, divisões políticas, isolinhas de temperatura).

Os dados no SR são organizados de uma maneira contínua e, através de análise quantitativa ou da fotointerpretação já discutidos anteriormente, esses dados podem ser classificados e agrupados em conjuntos discretos. Nesse processo ocorre o interfaceamento com os GIS.

A transferência de dados entre SR e um GIS raster é mais simples pois ambos os sistemas utilizam-se do pixel para o armazenamento. Entretanto, no interfaceamento entre o SR e os GIS vector a transformação de pixel em vetores geralmente é acompanhada por uma perda de informação. Esse fato acontece em função do pixel não poder delimitar tão precisamente quanto o vector em um espaço cartesiano.

A enorme geração de dados produzida por SR têm de ser armazenada, recuperada, manipulada, analisada e mostrada. A **tabela 3** mostra algumas das operações que os GIS podem fazer a respeito do manejo de uma base de dados.

Star (1990) aponta que uma área-chave para aplicações conjunta entre o SR e os GIS é a identificação de mudanças. O SR funciona como uma tecnologia para detectar a mudança enquanto o GIS é uma tecnologia para a quantificação da mudança.

A detecção de mudança pode ser feita comparando-se mapas analógicos com imagens ou com uma série multi-temporal de imagens. A

correção geométrica entre as imagens tem que ser feita da maneira mais precisa possível pois a comparação ocorre ao nível de pixel que representa alguma característica no solo.

| |
|---|
| Intersecção e sobreposição do conjunto de dados (máscara) |
| Intersecção e sobreposição de polígonos com dados espaciais |
| Identificação de formas |
| Identificação de pontos em polígonos |
| Determinação de áreas |
| Determinação de distâncias |
| Mapeamento temático |
| Cálculos de proximidade (rota mais curta) |
| Procura por localização |
| Procura por atributo definido pelo usuário |
| Procura por similaridade |

Tabela 3: Operações de manipulação de dados pelos GIS

Fonte: Richards (1993)

Star (1990) propõe, tendo em vista a série de dificuldades em se comparar o valor do brilho pixel a pixel, uma detecção de mudança simbólica. A primeira atitude do analista é decidir um conjunto de categorias temáticas que são importantes para a aplicação em questão. A comparação final é feita nas imagens já classificadas gerando tabelas estatísticas da mudança e uma matriz das probabilidades de mudança de uma categoria para outra (Flamm et al., 1994).

Aplicações e Vantagens

Contendo características multispectrais o SR é uma ferramenta excelente para observação e avaliação de características biofísicas dos recursos da terra. Como foi discutido anteriormente estas características permitem medir informações não discerníveis pelos olhos humanos.

Uma vantagem do SR frequentemente reivindicada é que dados remotamente adquiridos podem ser coletados em escalas multi-temporais e espaciais tornando-os extremamente útil para estudos de séries temporais.

Grandes áreas são cobertas periodicamente através de órbitas de satélite e os dados ambientais podem ser coletados de uma maneira abundante e constante. Medidas podem ser feitas repetidamente em uma escala regional e até mesmo global. Como se discutiu antes, há uma variedade de faixas espectrais e sensores acessíveis para o provimento desses dados.

Em países em desenvolvimento onde dados censitários e de levantamentos são escassos, o SR pode fazer uma contribuição significativa provendo informação atualizada tanto para os usuários do setor público como para os do privado.

O SR tem sido amplamente usado para pesquisa ecológica, planejamento regional e manejo ambiental. Quattrochi et al. (1994) indicam que o SR é uma excelente fonte de dados para validar modelos que prevêem modificação da paisagem. Muitos autores argumentam que o SR vem unificar as áreas da pesquisa ecológica e a dos projetos de avaliação de terra.

Basicamente quatro tipos de investigação ecológica podem usar técnicas de SR efetivamente: 1. inventário e mapeamento de recursos; 2. medição de atributos ambientais; 3. processos de delineamento de ecossistemas; e 4. análise de mudança e promoção de alternativas para melhor alocação de terra.

Como um campo de estudo próprio, o SR evoluiu significativamente desde o começo através da fotogrametria até a era de satélites de coleta de dados. O Processamento Digital de Imagens é um campo paralelo de investigação que também contribuiu muito para este processo evolutivo tornando disponível várias metodologias, técnicas e conceitos novos.

Todos os anos novos dispositivos de SR são lançados com desempenhos melhorados como a resolução espacial e uma melhor resolução radiométrica (por exemplo dados de radar). Este passo promoverá o desenvolvimento de mais aplicações e a melhoria dos usos atuais.

Para países em desenvolvimento estas novas perspectivas são extremamente importantes pela carência crônica de dados ambientais. O preço de aquisição deste tipo de dados e as dificuldades tecnológicas de manipulação ainda são fatores que ocasionam acentuada desvantagem em relação a outros métodos.

2.7 Redes Neurais Artificiais

2.7.1 Redes Neurais

A Tecnologia de Neurocomputação ou meramente, Redes Neurais Artificiais (RNA) é uma área das ciências da informação que adquire seus arquétipos inspirada pelos sistemas nervosos biológicos. Uma RNA é composta de muitos elementos simples (neurônios) trabalhando em paralelo imitando os neurônios biológicos. As funções na rede são definidas principalmente pela associação entre seus elementos. A RNA pode ser treinada ajustando os valores das sinapses (conexões) entre os muitos elementos.

O desenvolvimento do computador contribuiu significativamente para o interesse atual nas RNA. A capacidade aumentada de armazenamento e de executar tarefas intensivas tem permitido aos investigadores trabalhar com conjuntos de dados maiores (Clothiaux et al., 1994).

Diferentemente de computadores digitais que trabalham com algoritmos numéricos de alta precisão, nas RNA as funções são incluídas dentro da rede. Toda operação básica de computação tem um neurônio ou uma conexão para ser executada (Pao, 1989).

Problemas que foram resolvidos através da inteligência humana são agora possivelmente resolvidos pelas RNA por meio de procedimentos que têm semelhanças com os processos mentais humanos (aprendizagem, memória).

Openshaw (1994a) cita que três revoluções científicas aconteceram: a estatística (anos sessenta); a modelagem matemática (anos setenta); e a neurocomputação (nos anos noventa) – e esta última promoverá uma terceira

revolução quantitativa no campo da Análise Espacial (por conseguinte em GIS).

Ele discute que "...melhoria no desempenho e na eficiência, menos suposições críticas, facilidade de manipulação, expansões de aplicabilidade e eventualmente automatização..." tornarão as RNA uma ferramenta de análise quantitativa nova de extrema importância.

2.7.1.1 Função e Arquitetura

O modelo de rede neural de um único neurônio é apresentado na **figura 4**. Uma entrada de um escalar p é feita por uma conexão que multiplica seu poder pelo peso do escalar w , para produzir $w*p$, novamente um escalar.

No primeiro modelo a entrada ponderada $w*p$ é o único argumento para a função de transferência F (não há nenhuma influência) que produz a saída escalar a .

No segundo modelo o viés b é adicionada ao produto $w*p$. O viés age como um peso, um peso constante 1 de entrada. A soma da entrada ponderada $w*p$ e o viés b é a entrada da função de transferência n , um

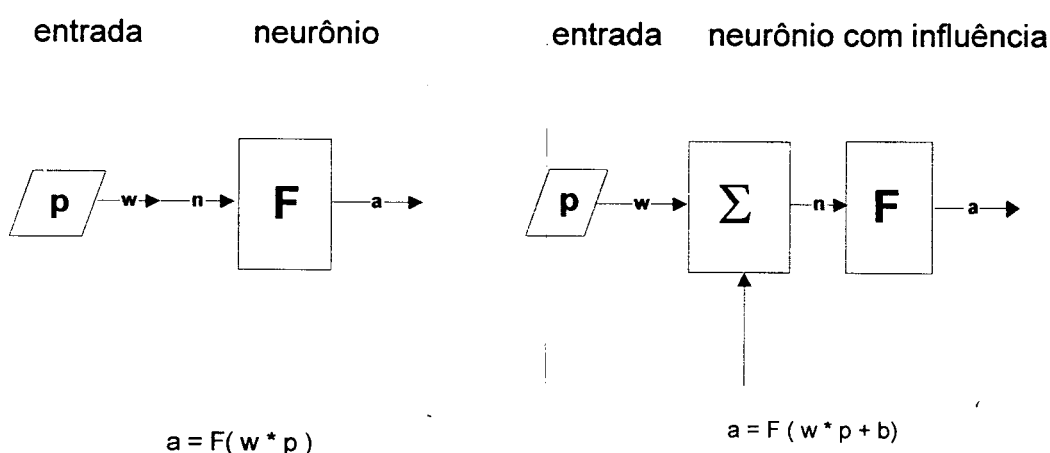


Figura 4: Modelo de Rede Neural de um único Neurônio com uma entrada

Fonte: Demuth, H., & Beale M. (1993) Neural Network TOOLBOX , The Math Works Inc.

escalar. Esta soma é o argumento da função de transferência F , tipicamente uma função de passo ou uma função sigmóide, que leva ao argumento n e produz a saída a .

Os escalares w e b são ambos parâmetros ajustáveis do neurônio. A idéia central das redes neuronais é que tais parâmetros podem ser ajustados de forma que as redes exibam algumas condutas desejadas ou interessantes (i.e. relações de causa-efeito). Uma rede pode ser treinada para fazer uma tarefa particular ajustando seus parâmetros (pesos e influência), ou por outro lado a rede ajusta estes parâmetros para alcançar algum fim desejado.

O viés b é um parâmetro ajustável do neurônio, ele não é uma entrada. Entretanto, a constante "1" que dirige a influência é uma entrada e deve ser tratada como tal.

Um exemplo de várias entradas para um único neurônio é apresentada em **figura 5**. As entradas individuais $p(1), p(2), \dots, p(R)$ são ponderadas pelos elementos $w(1,1), w(1,2), \dots, w(1,R)$ e os valores resultantes ponderados são entradas para a junção somatória. Essa somatória é simplesmente o produto do vetor linha w e o vetor coluna p .

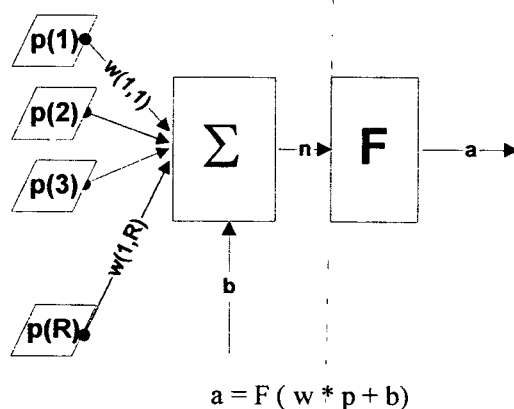


Figura 5: Modelo de Rede Neural de um único Neurônio com várias entradas
Fonte: Demuth, H., & Beale M. (1993) Neural Network TOOLBOX , The Math Works Inc.

A **figura 6** exemplifica uma rede neural com uma camada de neurônios. Nesta rede, cada elemento do vetor entrada p está conectado a

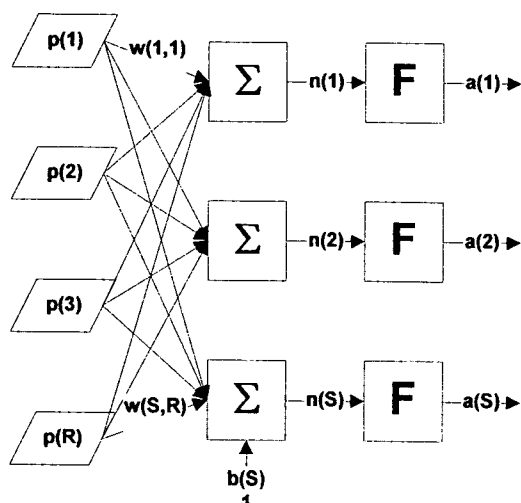


Figura 6: Modelo de Rede Neural com vários Neurônios e várias entradas

Fonte: Demuth, H., & Beale M. (1993) Neural Network TOOLBOX The Math Works Inc.

cada neurônio de entrada através da matriz de peso W . Muitos autores reconhecem que o conceito de RNA tem muitas analogias com o mundo estatístico. Um das possibilidades de aplicação da rede neural é considerá-la uma "super-forma de regressão múltipla" (Hewitson, 1994).

A rede tem a habilidade para achar qualquer função não-linear arbitrária que desenvolve uma relação $\{y\} = f\{x\}$ como em uma regressão, nesse caso não ficando condicionada a uma relação linear.

Há algum consenso em termos de sua eficiência melhorada para análise quantitativa; habilidade para conduzir com problemas de não-linearidade, e finalmente, possibilidades de automatização (Openshaw, 1994).

2.7.2 Modelagem Estatística

Capturar o padrão geral de um relacionamento é um dos objetivos principais da modelagem estatística (Smith, 1994). As técnicas de modelagem estatísticas são usadas para deduzir tais padrões derivando as equações a partir dos dados.

A equação derivada a partir dos dados pode ser interpretada como um mapeamento porque permite traçar qualquer ponto de variáveis independentes e dependentes no espaço de ambas.

Com uma variável independente há um mapeamento linear. Com duas variáveis independentes a equação de regressão define um plano em um espaço tridimensional. Uma equação de regressão multivariada define um espaço multidimensional no qual a função de mapeamento não pode ser visualizada geometricamente.

Essa função de mapeamento produz erros condicionando valores diferentes para a variável dependente. Um modo de estimativa do erro, conforme definido na seção 2.3.2, é medir a diferença média entre a estimativa e o resultado atual. Na verdade, matematicamente o quadrado das diferenças. A equação para esta medida de erro é

$$\mathcal{E} = \frac{\frac{1}{2} \sum_{n=1}^N (y_n - t_n)^2}{N} \quad (10)$$

em que \mathcal{E} é o erro, N é o número de exemplos para os quais o erro é medido, y_n é a previsão da equação para o exemplo n e t_n é o resultado verdadeiro para o exemplo n .

Existem algumas fontes de erros. O primeiro é o ruído que inclui imprecisões na medição dos dados e a escolha imprópria de variável independente. O segundo é que a função de mapeamento pode não representar o verdadeiro relacionamento presente nos exemplos.

A precisão de regressão linear pode ser limitada por causa de sua forma linear. Transformando as variáveis para tornar os dados lineares é uma forma de capturar relações não-lineares, mas várias dificuldades existem quando se pretende realizar tal meta.

O caminho mais prático é determinar uma função de mapeamento cuja forma seja construída pelos dados. A regressão logística e a polinomial são técnicas que vêm sendo extensamente usadas com este objetivo. Uma

técnica que tem sido usada, em que supostamente a equação pode tomar qualquer forma que os dados possuem, será discutida na próxima seção.

2.7.3 Backpropagation

Em 1974, Paul Werbos desenvolveu, em sua tese de doutoramento em Havard, uma rede neural supervisionada “feedforward” – o algoritmo de treinamento “backpropagation”. Infelizmente o seu trabalho permaneceu quase desconhecido para a comunidade científica.

Somente em 1982 Parker (em Rumelhart et al., 1986) redescobriu a técnica e em 1985 publicou-a em um relatório do M.I.T. Aproximadamente ao mesmo tempo Yan Le Cun (1986) também redescobriu o algoritmo (em Rumelhart et al., 1986). Mas talvez a melhor primeira descrição é a de Rumelhart, Hinton e Williams (1988). Este último trabalho tornou o algoritmo largamente conhecido.

Normalmente redes de “backpropagation” têm uma ou mais camadas escondidas de neurônios seguidas por uma camada de saída de neurônios. Camadas múltiplas de neurônios com funções de transferência não-lineares permitem à rede aprender relacionamentos não-lineares e lineares entre vetores de entrada e de saída.

As únicas conexões permitidas estão entre os nós em camadas consecutivas e dirigidas para frente. Por esta razão é que o neurônio de multicamadas é chamado uma rede de “feedforward”.

Quando uma rede de “backpropagation” é inicialmente alimentada com um padrão ela faz uma distribuição aleatória de seus pesos para o que aquele padrão possa parecer. Compara então quão distante suas respostas foram da resposta real fazendo um ajuste de seus pesos de conexão.

A regra de aprendizagem usada na “backpropagation” é a regra delta. A aprendizagem é um processo supervisionado que acontece com cada ciclo ou época (cada vez que se apresenta à rede com um novo padrão de entrada) através de um fluxo para frente de saídas e propagação reversa de erros de ajustes de peso.

A rede “backpropagation” executa uma descida de gradiente dentro do espaço de solução do vetor para um mínimo global ao longo da superfície do erro. Este mínimo global é a solução teórica com o erro mais baixo possível.

Na maioria dos problemas o espaço de solução é bastante irregular o que pode causar um estancamento da rede em um mínimo local que não é a solução global pretendida.

Uma vez que uma rede é treinada, ela pode ser usada como uma ferramenta analítica para dados novos (dados não vistos). O usuário somente permite à rede trabalhar em modo de propagação para frente. Novas entradas são apresentadas e são processadas pelas camadas internas não acontecendo aqui a propagação reversa. A saída de uma corrida para a frente é o modelo previsto para os dados que podem ser usados para análise posterior.

As RNA são aproximadores universais e devem ser usadas onde alta tolerância de erro é possível. Vários trabalhos podem ser feitos que se incluem dentro destas restrições, tais como, captura de associações ou descoberta de regularidades dentro de um conjunto de padrões; quando existe um número de variáveis e diversidade de dados muito grande; quando a relação entre variáveis é vagamente compreendida; quando ocorrem relacionamentos de difícil descrição através de aproximações convencionais.

Existem algumas limitações. As redes neurais de “backpropagation” são de certo modo como caixas pretas. As únicas coisas que podem ser controladas são: definição da arquitetura geral; semente inicial com números aleatórios e espera pela rede produzir uma saída desejada. O produto final é uma rede treinada que não provê equações ou coeficientes como na regressão que define um relacionamento.

O fato da maioria dos computadores possuírem arquitetura serial pode ser um fator extremamente consumidor de tempo, o que pode tornar-se problemático com uma grande quantidade de dados e em redes neurais muito grandes.

A regra delta generalizada, um enfoque de treinamento supervisionado, é usado para mapear os dados (Windrow et al., 1960). Esta

regra foi generalizada da regra delta (eq. 3) para conter uma ou mais camadas de neurônios escondidos.

$$\mathbf{W}(k) = \mathbf{W}(k-1) + \eta [\mathbf{t}(k) - \mathbf{W}(k-1) \mathbf{x}(k)] \mathbf{x}^T(k) \quad (11)$$

em que $\mathbf{x}(k)$ é o padrão de entrada, $\mathbf{t}(k)$ é a saída desejada, $\mathbf{W}(k)$ é o estado da matriz de pesos que descreve a rede depois de k apresentações, e η é uma taxa de aprendizagem.

A regra delta é semelhante ao método matemático de aproximação estocástica para problemas de regressão. Porém, esta regra não assegura a discriminação dos dados que não são linearmente separáveis (como dados de fontes diversas) e uma generalização é feita para incluir uma ou mais camadas escondidas. Esta propriedade de generalização permite treinar uma RNA em um conjunto representativo de pares de entrada/saída sem a necessidade de treinar todos os possíveis pares de entrada/saída.

3. Desenvolvimento do Modelo

3.1 Introdução

A natureza básica da percepção humana é que nossa mente pode organizar, a variedade de estímulos que recebe, e ordenar esta vasta quantidade de dados crus e caóticos em informação útil. De uma maneira geral, o trabalho de percepção é dividido entre os vários processos sensoriais para obtenção desses dados crus e a faculdade para converter estes dados em informação significativa, experiência do mundo real.

A procura por uma estrutura de representação ótima para os dados ambientais tem vários problemas, os dois mais importantes a solucionar são: a relevância dos subconjuntos dentro da vasta quantia de dados tem que ser determinada e uma estrutura representacional tem que ser desenvolvida (Chalmers et al., em Hofstader, 1994).

O desenvolvimento de um modelo para a representação do padrão da paisagem faz uma analogia com este modelo de *percepção-representação* presente na mente humana. Uma integração entre o Sensoriamento Remoto (SR), os Sistemas de Informação Geográfica (GIS) e as Redes Neurais Artificiais (RNA) formam uma base conceitual e prática para desenvolver um modelo digital de percepção-representação para simulação de mudanças no padrão da paisagem.

O SR tem sido usado como uma fonte de dados ambientais varrendo a superfície de terra e obtendo grandes quantidades de dados - *percepção*. Os GIS têm a capacidade de manipulação, com uma estrutura para transformar e utilizar estes dados em vários formatos e fazer uso de outras formas de dados auxiliares tais como as áreas de censo, os levantamentos de campo, os sistemas de posicionamento global (GPS) - *representação*. As RNA podem encontrar a relevância dos subconjuntos de dados (padrões) estabelecendo relações entre tipos e formatos diferentes de dados (dados multivariados) - *percepção e representação*.

É comumente aceito o conceito que a interação entre processos ecológicos e as atividades humanas formam padrões espaciais que podem ser inferidos em uma determinada escala com consequente delimitação da área de ocorrência. Uma manifestação típica desta interação entre processos naturais e atividade humana é o padrão de uso da terra (ou cobertura da terra) observado em uma área (Lee et al.,1992).

As evidências têm mostrado que só algumas poucas variáveis-chaves podem explicar as interações entre os processos que acontecem ao nível da paisagem. Esta afirmação obviamente depende da escala espacial e temporal levada em consideração.

Se estas variáveis-chaves puderem ser reconhecidas e mapeadas em um sistema de representação e suas possíveis relações de causa-efeito entendidas dentro desse sistema, então, modificações ambientais podem ser previstas por meio de padrões presentes nos dados organizados (informação).

Lee et al. (1992) citam que processos ecológicos, forças de mercado e fatores sociais são os fatores motrizes para as práticas de uso do solo que

resultam na formação dos padrões da paisagem. Uma equação pode ser formulada que mapeia o padrão da paisagem como uma função destas forças motrizes:

$$\text{Padrão da paisagem} = F \left(\begin{array}{l} \text{processos ecológicos,} \\ \text{forças de mercado,} \\ \text{fatores sociais} \end{array} \right) \quad (12)$$

Estes autores argumentam que a estrutura da paisagem é uma função do uso da terra e que este uso é influenciado por estas três forças motrizes. O padrão de uso da terra ou de cobertura da terra é uma "manifestação" das relações entre os vários processos ecológicos e atividades humanas sob a superfície de uma região.

Nesse ponto as questões levantadas por Flamm e Turner (1992) são cruciais sob o ponto de vista do desenvolvimento de um modelo de avaliação de modificações no padrão de paisagem.

Primeiramente, os autores indagam se a informação derivada das disciplinas sócio-econômicas e biológicas podem ser representadas e integradas como uma estrutura comum de dados.

Em segundo lugar questionam se poderia haver uma generalização para os problemas ambientais de manejo e conservação da paisagem.

E, integrando as duas questões, seria possível que uma solução única, através do enfoque de um modelo de simulação, pudesse alcançar todos os aspectos de manejo e conservação das paisagens?

3.2 Construção do Modelo

3.2.1 Antecedentes

A integração entre o SR e os GIS desenvolveu-se fortemente nas últimas duas décadas. O SR tem sido especialmente utilizado como uma constante fonte de dados para os GIS, especialmente em países em desenvolvimento (ou emergentes) onde a carência de dados é aguda.

Com o advento de satélites para observação da Terra, uma quantidade volumosa de dados tem sido coletada desde os anos setenta. Somado a esses dados, temos os dados de origem censitária; os de levantamentos do solo; os de aerofotogrametria e outras fontes, o que ocasiona um desafio enorme de interpretação e transformação em informação útil para a gestão ambiental.

Mais recentemente as RNA (a partir de meados da década de oitenta) têm sido usadas para classificar dados adquiridos por dispositivos de SR em classes de uso do solo. Algumas tentativas de integração das RNA como um módulo operacional analítico dos GIS também foram feitas.

Silveira et al. (1996) indicou que uma RNA treinada com um conjunto de dados censitários, variáveis ambientais e imagens de SR pode ser usada para estabelecer associações e prováveis modificações no padrão da paisagem.

Lo (1996) faz uma aproximação integrada entre variáveis socio-econômicas e biofísicas e conclui que na ausência de dados censitários atualizados, imagens de satélite de alta-resolução podem ser empregadas para avaliar a condição do desenvolvimento sócio-econômico em uma região.

Yamamoto (1996) desenvolveu um modelo de avaliação de terra onde uma RNA treinada pode expressar as relações causais mais objetivamente quando comparada com avaliações ambientais convencionais.

Flamm e Turner (1992) apresentam um modelo desenvolvido para a região da bacia do Rio LittleTennessee no Estado norte-americano da Carolina do Norte. O modelo denomina-se LUCAS (Land Use Change Analysis System) e foi desenvolvido para estimar as consequências de ações alternativas de manejo e conservação da paisagem.

Eastman (1995) indica que o monitoramento de mudanças na paisagem tem novas perspectivas com a evolução dos GIS e do SR. Apesar desse otimismo, técnicas de análise de séries temporais e de mudança não estão bem desenvolvidas.

A detecção do fenômeno de mudanças na paisagem não é de solução trivial. Apesar das modificações do padrão da paisagem serem facilmente

medidas, o problema maior diz respeito à diferenciação da mudança real (induzida pelas atividades humanas) daquela induzida pela variabilidade existente no meio ambiente e até mesmo daquela produzida pelo processo de amostragem.

3.2.2 Desenvolvimento

A divisão do espaço físico em pacotes de amostragem homogêneos (Openshaw 1993) permitiu a aplicação da matemática e estatística convencionais em análise espacial pelo uso de mapas que representam valores em um espaço numérico.

Os GIS permitiram a mudança de imagens de mapas analógicos para dados digitais que expressam a quantidade de variáveis físicas, sociais ou econômicas. Os dados agora podem ser manuseados como mapas e os mapas como variáveis. Os mapas tornaram-se um arranjo de números geograficamente referenciados.

Estes mapas podem ser usados como variáveis quantitativas ou qualitativas únicas que executam as mesmas operações estatísticas ou matemáticas básicas gerando mapas derivados. Operações algébricas gerais podem ser conduzidas como um conjunto de imagem de mapas como variáveis ao invés de números.

Desta maneira, variáveis com uma referência geográfica (localização na superfície da terra) podem ser espacialmente distribuídas e representadas, em um ambiente de GIS, como mapas raster ou vetor.

De uma maneira geral, o que a estatística faz é reduzir a complexidade dos dados para um valor distinto. Com os mapas de GIS, esta complexidade é mantida descrevendo-se espacialmente a diferença dos dados.

Como foi discutido na seção 2.5.3, uma imagem raster é uma malha fina de células no qual uma condição ou característica (ou atributo) da superfície da terra naquele ponto é registrada. Os dados de uma imagem digital são compostos de sequências bidimensionais (colunas e linhas) de elementos de quadro discretos (pixels). Esta imagem é uma matriz numérica

e cada pixel (elemento de quadro) representa o valor de atributo de cada elemento desta matriz.

Os GIS têm a capacidade de sobreposição de imagens obtendo imagens compostas adicionando multidimensionalidade vetorial ao pixel. Neste momento um elemento fundamental de informação espacial $T = \langle x, y, z_1, z_2, \dots, z_n \rangle$ pode ser definido, que representa os valores de n variáveis de espaço (z) na localização (x, y) (Goodchild, 1992).

O autor argumenta que o fato de T representar o elemento fundamental de informação espacial causa algumas suposições necessárias para tornar esta modelagem factível: cada z é estimado como único para qualquer localização; T pode ser confirmado experimentalmente. Estas variáveis podem ser de qualquer tipo e formato de dados tais como binárias, nominais, ordinais, de intervalo, ou razão.

Centrado na enorme capacidade de manipulação de dados dos GIS, um modelo é proposto para investigar a formação e a modificação do padrão da paisagem como uma função de possíveis variáveis que possam induzir esse processo (forças motrizes).

O propósito geral do modelo, como em uma regressão multivariada, é aprender mais sobre as possíveis associações entre as variáveis. A identificação do conjunto de variáveis que formam os padrões da paisagem, pode colaborar para exercícios de simulação sobre o uso da terra, e a averiguação das conseqüências prejudiciais que uma alocação errada dos recursos naturais pode causar. Esse propósito vem de encontro à necessidade do desenvolvimento de ferramentas para o auxílio na Avaliação de Impactos Ambientais (AIA - seção 2.2.2).

É importante distinguir neste ponto que o tipo de modelo escolhido é o descrito na seção 2.5.5 (**figura 3**), onde graus diferentes de integração de modelos e os GIS são descritos. Na integração proposta há somente uma troca de arquivos (informação). Basicamente existem modificações de formatos de arquivo e rotinas de entrada/saída. O GIS é usado para armazenar, manipular e gerar os dados de entrada do modelo e, mais tarde, exibir os resultados.

Dois problemas surgem quando uma estrutura para um modelo ambiental é proposta: O primeiro problema é que os fenômenos ambientais (incluindo processos ecológicos e sociais) normalmente não podem ser descritos por relações de causa-efeito simples e determinísticas. O segundo problema é que os dados disponíveis para este tipo de modelagem têm uma variedade de formatos diferentes e distribuições desconhecidas.

Não obstante, algumas aproximações podem ser feitas. A procura da solução para o problema é dirigida pelos dados. Em outras palavras, os dados devem indicar quais são os padrões prováveis de relacionamentos presentes em um lugar especificado.

A modelagem é uma típica regressão não-linear que computará o relacionamento entre um conjunto de variáveis independentes e uma variável dependente.

A rede não especifica a natureza da relação, como em uma regressão linear onde se assume que os relacionamentos são lineares por natureza. Relações lineares simples são muito convenientes e permitem fazer interpretações diretas como "o mais de x, ocasiona o mais proporcional de y", e "um determinado aumento particular em x, pode ser esperado um aumento proporcional em y". Normalmente relações de não-linearidade não podem ser interpretadas e verbalizadas de uma maneira tão simples.

O fato de, normalmente em problemas ambientais, qualquer tipo de função de mapeamento entre as variáveis independentes e variáveis dependentes ser permitido, levanta duas questões importantes. Primeiro, que tipos de relacionamentos têm algum significado, em outras palavras, eles são interpretáveis de alguma maneira? O segundo ponto é como exatamente quantificar a relação, ou como chegar a resultados que nos permitem dizer se ou não há uma relação de não-linearidade como prevista?

A estimativa de não-linearidade é uma operação de ajuste comum que estimará qualquer tipo de relacionamento entre uma variável dependente (ou variável de resposta) e uma lista de variáveis independentes. Em geral, todos os modelos de regressão podem ser declarados como:

$$y = F(x_1, x_2, \dots, x_n) \quad (13)$$

Em condições mais gerais, o que se procura entender é como uma variável dependente (y) está relacionada a uma lista de variáveis independentes (x_n); o termo $F(x...)$ na expressão acima significa que y , a variável dependente ou de resposta, é uma função de x_n , ou seja, as variáveis independentes.

A estimativa não-linear permite especificar essencialmente qualquer tipo de modelo de regressão contínuo ou descontínuo. Quando o modelo de regressão linear simples não descreve suficientemente as relações entre variáveis, então uma aproximação pelo modelo de regressão não-linear é a mais apropriada.

A expressão dados de fontes diversas ("multisource data") é uma denominação para vários tipos e formatos de dados com origens diversas, relativos à mesma cena de estudo.

Vários problemas influenciam um modelo que use estes tipos de dados através de uma análise estatística multivariada convencional. Estes problemas são, formatos dos dados; unidades de medida; escala; imprecisão de amostragem; distribuições estatísticas diferentes; peso dos dados em problemas de classificação.

As RNA são apropriadas para conduzir, como em uma "modelagem estatística alternativa", esse tipo de problema com tal variedade de dados. O processo a ser modelado pela RNA é semelhante a uma regressão multivariada como discutido na seção 2.7.2. A RNA pode determinar quanto significado cada conjunto de dados tem no modelo. Os dados "decidem" qual forma a função de mapeamento deve ter. A função de mapeamento é complexa o suficiente para ser flexível e poder tomar qualquer forma que os dados requeiram.

A **figura 7** mostra o modelo proposto que é centrado em um GIS uma vez que utiliza as capacidades dos GIS para receber, gerar e preparar os dados para serem usados como as variáveis de entrada e saída para o

treinamento de uma RNA. Estas variáveis de entrada/saída são representadas por imagens raster de uma área ser avaliada.

3.3 Representação dos Dados

3.3.1 Variáveis de Entrada

A equação proposta na seção 3.1 faz uma suposição de que três forças motrizes geram o padrão da paisagem e por conseguinte são os indutores da sua modificação:

Padrão da paisagem = $F(\text{processos ecológicos, forças de mercado, fatores sociais})$

Baseado nesta suposição, as variáveis *independentes* de interesse são preparadas oriundas de várias fontes: econômicas, sociais e ecológicas.

Considerando que a idéia principal do modelo é identificar o conjunto de possíveis variáveis-chaves que induzem a formação e a modificação do padrão da paisagem, elas são geradas por critérios diversos e, como numa regressão “stepwise”, o peso dos sub-conjuntos dessas variáveis no modelo é gradativamente avaliado.

A **figura 8** mostra as diversas possibilidades de geração das variáveis independentes, de acordo com os interesses de investigação.

Uma fonte muito importante destas variáveis são aquelas coletadas em bancos de dados das várias entidades governamentais (agências federais e estaduais) armazenados principalmente como dados vector de GIS. Estes dados são os mapas topográficos e temáticos que podem ser convertidos por procedimentos de GIS tais como *interpolação*, “*rasterização*”, *distância*, *atribuir* e outros para a geração dos mapas derivados (superfície contínua).

Por exemplo, no caso de procedimentos de *distância*, o GIS calcula a distância a partir de uma característica ou de um conjunto de características. Na imagem resultante, a todo pixel é atribuído um valor que representa sua

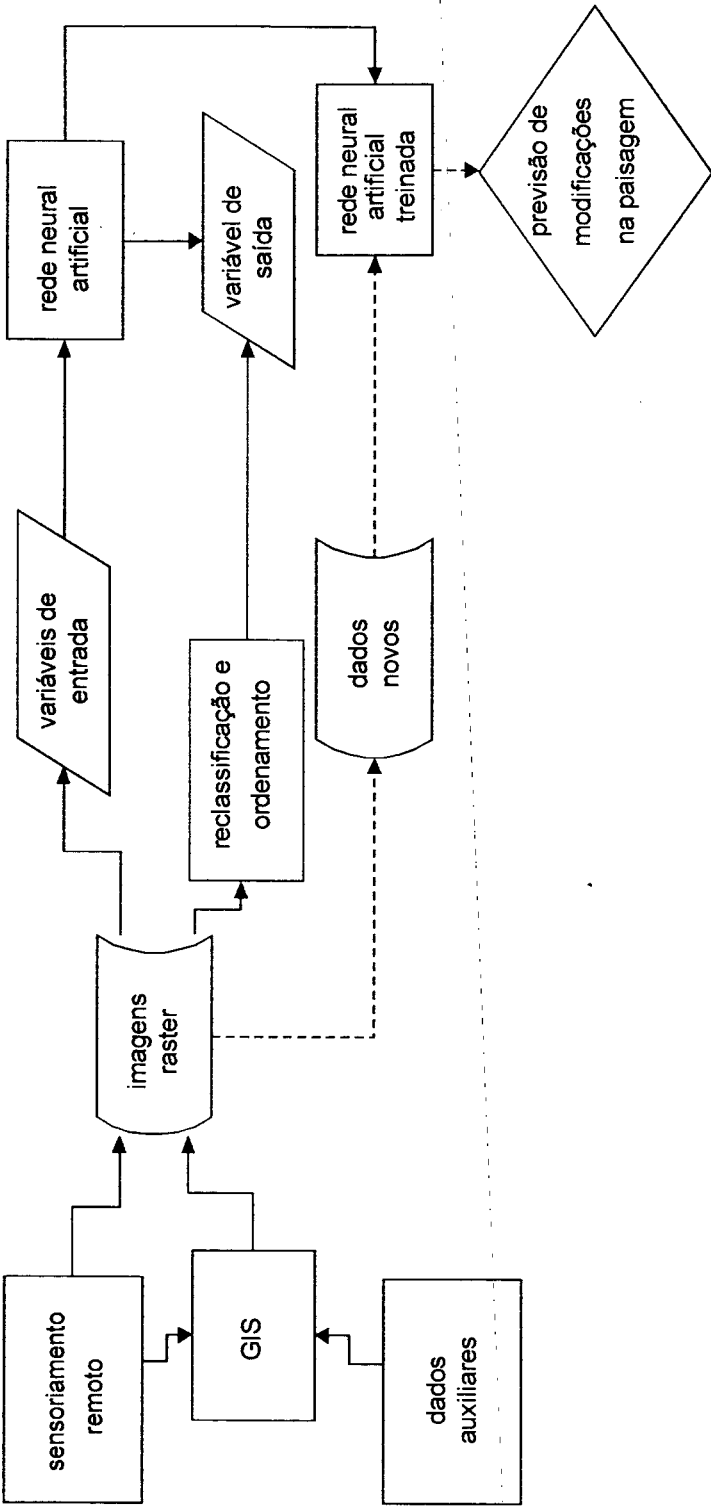


Figura 7: Modelo baseado na Integração de SR, GIS e RNA

distância da característica mais próxima. Proximidade a sedes de distritos rurais, rios e estradas são usadas como variáveis *independentes* uma vez que esse tipo de variável provavelmente tem uma influência social e econômica muito forte na formação do padrão da paisagem.

As características da terra (hidrografia, topografia, vegetação) representadas no GIS podem ser associadas com os atributos presentes em um banco de dados relacional (Sistema de Administração de Banco de dados - DBMS). Quando usado em associação com um GIS, um DBMS permite que a informação presente no banco de dados se relacione com os arquivos vector ou raster de definição geográfica. Tendo uma referência espacial, essas características podem ser distribuídas de acordo com uma característica geográfica relacionada.

Uma vez que a maioria de dados socio-econômicos não estão associados a qualquer tipo de característica geográfica, uma solução possível é associar estes dados às divisões políticas municipais para que se possa distribuir espacialmente estas variáveis. Parte da **figura 8** mostra como esta distribuição espacial é levada a cabo.

A partir de um plano vetorial da divisão municipal é possível gerar uma imagem contendo uma variável de interesse dentro da área de localização do município. Para cada pixel dentro de um limite administrativo municipal (divisão municipal) é nomeado um atributo único dependendo do tipo de variável independente que está sob análise. Através do procedimento *atribuir (assign)* em um ambiente de GIS pode-se exibir os atributos de interesse (dados censitários) presentes no banco de dados relacional para um município específico.

3.3.2 Variável de Saída

O fato da RNA mimetizar um “espécie de super-forma de regressão multivariada” torna a escolha da variável de saída, ou supostamente, a variável resposta (dependente) um procedimento importante, pois a

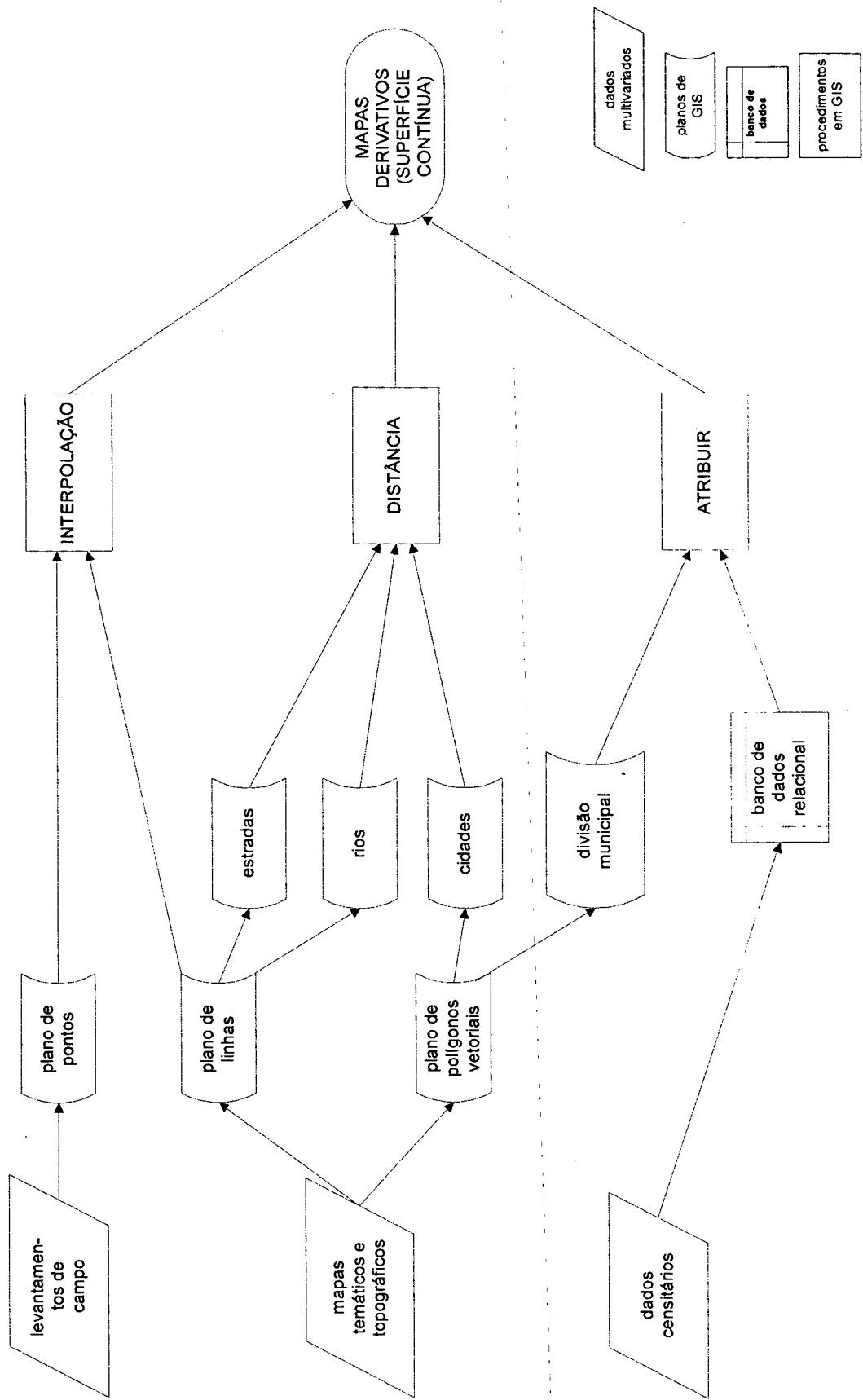


Figura 8: Geração das variáveis independentes

sensibilidade do modelo é indicada pela variável resposta. Essa variável de saída deve representar um padrão de mudança na paisagem.

Conforme discutido na seção 3.2.1 a avaliação das mudanças na paisagem não é um processo trivial. Geralmente a mudança indica não somente diferença na cobertura do solo entre duas datas mas também a diferença que advém fora da variação normal (variabilidade ambiental).

A análise de séries temporais examina a variação de uma sequência de imagens enquanto que a análise de mudança é o exame de modificações entre duas datas. Resumidamente a primeira diz respeito à análise de comparações múltiplas e a segunda entre duas imagens.

Eastman (1995) elenca uma série de técnicas de avaliação de mudanças e séries temporais no padrão da paisagem.

Inicialmente essas técnicas são divididas em comparações em pares de imagens e comparações entre múltiplas imagens. No primeiro grupo encontramos as análises quantitativas de diferenciação; de razão; de regressão de imagens e a análise do vetor de mudança. Também nesse grupo, somente que análise qualitativa, encontra-se a classificação cruzada (crossclassification).

No segundo grupo as análises são feitas através de comparações entre uma série múltipla de imagens. Elas são: desvio padrão; vetor de mudança; sequenciamento e perfil temporal; e a análise dos principais componentes.

Na escolha da técnica mais apropriada deve-se considerar primordialmente o objetivo a que o trabalho se propõe a resolver, mas também considerar a disponibilidade dos dados para análise.

Com essas limitações em mente a Análise do Vetor de Mudança mostrou-se como a mais representativa para a geração da variável de saída dentro do modelo proposto.

Nessa análise as bandas de cada data da imagem de satélite compreendem um espaço multidimensional para cada pixel. Uma vez que os pixels tem uma localização nesse espaço de várias bandas, a diferenciação entre todos os pixels na mesma localização, em duas datas distintas, pode ser expressa pela sua distância Euclidiana neste espaço (**Figura 9**). A direção do vetor, expressa através de seu ângulo, descreve também o tipo de mudança que pode ter ocorrido.

A fórmula da distância Euclidiana:

$$D = \sqrt{(b1_{90} - b1_{94})^2 + (b2_{90} - b2_{94})^2} \quad (14)$$

é utilizada dentro de um ambiente de GIS para gerar uma imagem multidimensional da distância onde $b1$ e $b2$ representam as bandas de satélite 1 e 2, e, onde 90 e 94 representam os anos de 1990 e 1994 respectivamente.

Esse tipo de análise pode ser utilizado tanto para comparações entre duas imagens como para séries temporais de imagens. Conforme descrito na seção 4.2 as imagens de satélite disponíveis são dos anos de 1990 e de 1994 possibilitando somente uma análise comparativa entre duas datas.

Esse fato culminou em indicar a Análise do Vetor de Mudança como a determinante da variável de saída representativa da mudança no padrão da paisagem.

3.3.3 Amostragem dos Dados

O processo inteiro de construção das variáveis *dependente* e *independentes* tem por objetivo produzir vários mapas derivados, em formato raster (imagens), onde cada pixel tem exatamente a mesma localização (coordenadas geográficas) de seu correspondente em qualquer das imagens.

A concepção do elemento fundamental de informação espacial $T = \langle x, y, z_1, z_2, \dots, z_n \rangle$ - n variáveis espaciais (z) na localização (x, y) (Goodchild, 1992) apresentado na seção 3.1.2, pode ser realizada através da amostragem de várias imagens raster, representando n variáveis espaciais (z) na localização x, y e o processo de sobreposição de todas elas (variáveis *independentes*) pode ser conduzido mais tarde formando-se um vetor coluna adicionando multidimensionalidade espacial ao pixel.

Este artifício permitirá durante o procedimento de amostragem coletar dados multivariados exatamente de qualquer localização desejada simplesmente amostrando variáveis diferentes (imagens) e agrupando-as no vetor.

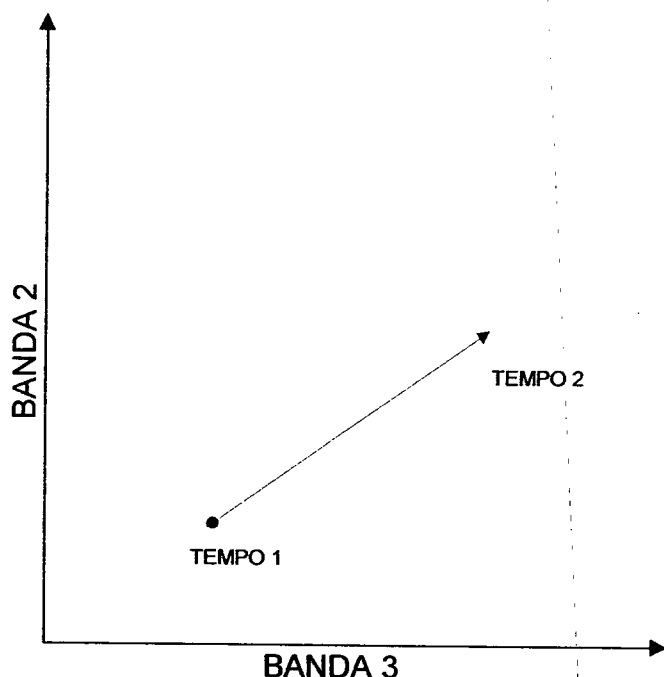


Figura 9 Distância Euclidiana no espaço

Esse processo permite suplantarmos também um limite de memória e da utilização da álgebra booleana nos GIS uma vez que estas operações são feitas em par a par. No modelo pode-se treinar a RNA com conjuntos de variáveis independentes ao invés de operações individuais entre pares de variáveis.

A **figura 10** mostra como o procedimento de amostragem é conduzido. Depois de gerar todas as variáveis de interesse no ambiente de GIS, um arquivo vector de localizações de pontos de amostragem é gerado para amostrar as variáveis dependente e independentes.

Este arquivo vector de pontos amostrados é obtido a partir de uma imagem raster que define a área de interesse a ser amostrada. Como um resultado final do processo de amostragem são produzidas estatísticas sumárias em um arquivo de valores de atributos.

Estas estatísticas resumidas no arquivo de valores indicam o mínimo, máximo, total, média, desvio padrão de todos os pontos na imagem analisada dentro de cada região definida em uma imagem de definição da característica (imagem de uma variável). Este arquivo está em formato ASCII com números exibidos em uma única coluna.

Neste momento, os dados na forma do arquivo de valores de atributos migram de um ambiente de GIS e vão para um ambiente matemático (MATLAB). Um aplicativo escrito na linguagem do Matlab transforma este arquivo coluna de valores em arquivos de linha e prepara um arquivo de entrada em um formato compatível para ser usado pela RNA.

De acordo com o interesse de treinamento da rede, cada variável independente pode ser progressivamente adicionada ou suprimida formando o vetor coluna que representa os dados de entrada de interesse para análise. Um processo semelhante à uma regressão multivariada é levada a cabo apresentando um conjunto (amostras das imagens de raster) de vetores de entrada e os vetores de saída correspondentes para a RNA. A RNA é treinada até que possa aproximar uma função de mapeamento que associe vetores de entrada e de saída.

O que é importante notar é que esta técnica permite superar uma limitação severa em operações de análise espacial. A manipulação de vários tipos e formatos de dados leva aos problemas estatísticos apontados anteriormente e especialmente às limitações de memória do computador.

Utilizando a RNA e fazendo uma regressão matemática com dados incluídos como vetores (cada vetor representando uma localização geográfica com vários atributos) torna possível dominar uma técnica praticamente irrestrita de manipulação de dados de fontes diferentes ("multisource data") com o objetivo de executar vários tipos de análises com dados múltiplos.

Não há distinção alguma entre os percentuais de variáveis qualitativas ou quantitativas a serem usados pelo modelo como variáveis independentes. Uma vez que a rede supostamente tem a habilidade para achar relacionamentos com qualquer tipo e distribuição de variáveis, não há uma preocupação com enfoques quantitativos ou qualitativos.

Depois de treinada a RNA pode ser usada para simular os relacionamentos entre as variáveis de entrada (*independentes*) com a de saída (*dependente*). A simulação conduzida com dados não vistos pela RNA pode gerar padrões diferentes da paisagem auxiliando na compreensão dos fenômenos de modificação da mesma. A formulação de diferentes combinações de variáveis independentes

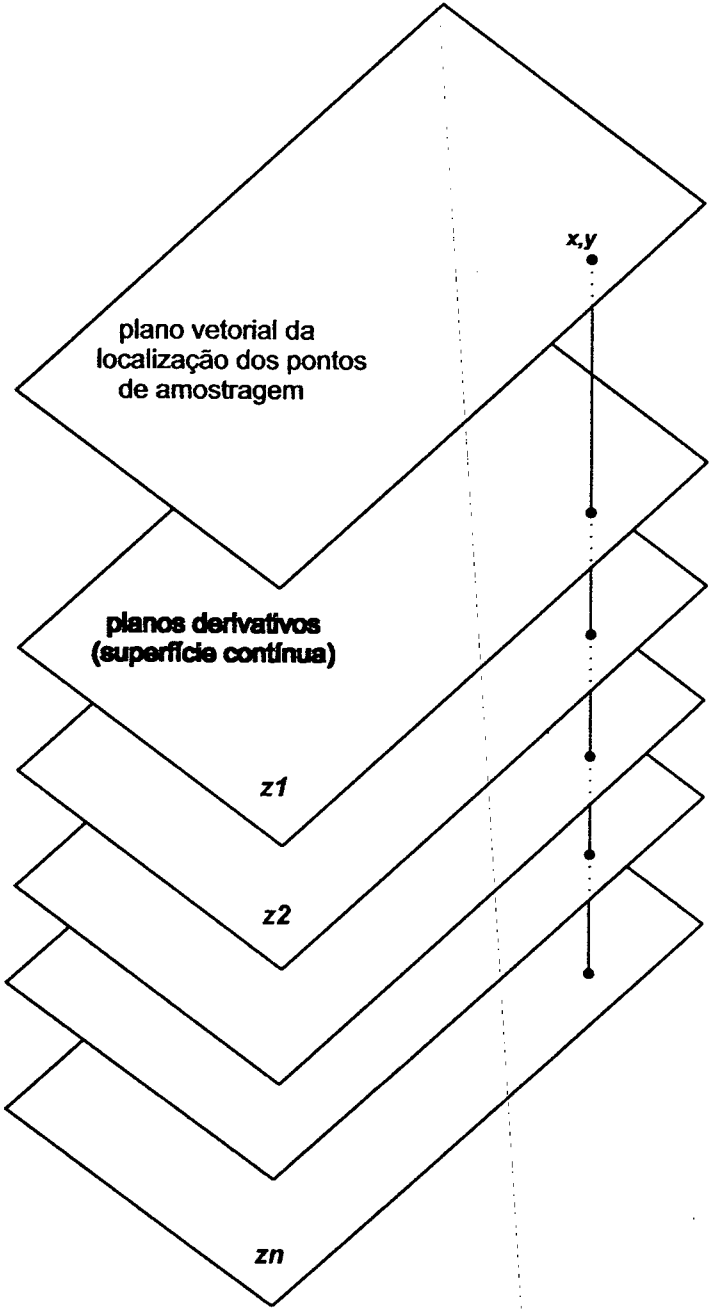


Figura 10: Procedimento de Amostragem

pode formar diferentes cenários de simulações para identificação das variáveis-chave que podem induzir à modificações na paisagem.

A RNA pode “aprender” a fazer previsões baseada em dados reais e, conforme discutido, esse modelo pode ser útil para identificação de variáveis-chave que induzem mudanças na paisagem.

4. Aplicação Prática

4.1 Caracterização da Área de Estudo

4.1.1 Aspectos Físicos

A área de estudo está situada em uma Região denominada Litoral Centro-Norte do Estado de Santa Catarina. A área de interesse está localizada nas coordenadas geográficas a 25° 57' 00" a 26° 30' 00" de latitude de Sul e 43° 30' 00" a 49° 20' 00" de longitude Oeste.

Esta área compreende a Bacia do Rio Cubatão e inclui toda a área da Cidade de Joinville e partes de outros municípios circunvizinhos.

A área do distrito municipal de Joinville é de 1.183 km² e os limites são ao Norte com os distritos municipais de Garuva e São Francisco do Sul; para o Sul com os distritos municipais de Araquari e Guaramirim; para o Leste com o distrito municipal de São Francisco do Sul e para Oeste com os distritos municipais de Jaraguá do Sul, Schroeder e Campo Alegre.

A *geologia* desta Região é caracterizada por duas áreas geológicas denominadas: A Área do Escudo do Atlântico e da Área de Sedimentos Quaternários.

O domínio do Escudo do Atlântico é composto de vários complexos constituídos por uma assembléia de rochas metamórficas, principalmente da fácies granulito, composição básica intermediária, geradas no Arqueano e Proterozóico Inferior e uma assembléia de rochas gnáissicas, graníticas e cataclásticas geradas no Proterozóico Superior (antes de 1.000 anos)

O domínio de Sedimentos Quaternários corresponde aos depósitos sedimentares não consolidados presentes perto do litoral, formados em ambientes

marinhos, fluviais, eólicos, lagunares ou misturados, durante o Holoceno. Estes depósitos consistem em areias, argilas, cascalhos, seixos e sedimentos siltico-argilosos.

A geomorfologia da Região é dividida em duas unidades geomorfológicas: as Planícies Litorâneas e a Serra do Mar.

A primeira corresponde a uma faixa estreita encontrada na porção oriental do Estado perto do Oceano Atlântico, onde praias arenosas e dunas mostram a evidência de processos eólicos e marinhos.

Ao longo de sua extensão, a geomorfologia da unidade é diversificada com a presença de penínsulas e baías entre as quais desenvolveram-se planícies arenosas descontínuas que abrigam incontáveis praias.

As altitudes médias nas Planícies Litorâneas são ao redor de 10m e alcançam, em terraços mais interiores, nas proximidades das montanhas e no oeste até 30m de altitude.

Os sambaquis, depósitos de origem humana de conchas, existem ao longo da costa do Estado de Santa Catarina e sua presença atesta a existência de um nível marinho mais alto.

A segunda unidade geomorfológica é denominada Serra do Mar. Constitui-se de uma prolongação da escarpa sul do Planalto de São Paulo, conhecido regionalmente como Serra do Mar.

A Serra do Mar em Santa Catarina somente aparece no extremo nordeste do Estado e seu limite meridional coincide aproximadamente com o distrito municipal de Joinville. Ela apresenta-se como um grupo de cristas e picos, separados por vales profundos em " V " e com escarpas íngremes. Nesta unidade geomorfológica alguns picos alcançam 1.500m de altitude.

A hidrografia do Estado de Santa Catarina é representada por dois sistemas independentes de drenagem - o sistema integrado da vertente do interior, comandado pela Bacia dos rios Paraná-Uruguai e o sistema da vertente Atlântica, formados por um grupo de bacias isoladas.

A Serra Geral é o grande divisor das águas que escoam para a Bacia do Rio Uruguai e os que vão para leste e escoam diretamente no Oceano Atlântico. Mais

ao Norte, a Serra do Mar serve como um divisor entre a Bacia do Rio Iguaçu e as bacias da vertente Atlântica que drenam a costa norte.

O sistema de drenagem da vertente Atlântica aonde a Região de Joinville está localizada, compreende uma área de aproximadamente 35.298 km², correspondendo a 37,00% da área total do Estado. A Bacia do Rio Cubatão tem 472 km² de área.

As bacias hidrográficas do Estado de Santa Catarina contam com densidade de drenagem (relação do somatório do comprimentos dos cursos de água de uma bacia pela sua área total) relativamente alta (0,5 km/ km² para drenagem pobre e 3,5 km/ km² para bacias bem drenadas).

A bacia do rio Cubatão (Norte) apresenta uma densidade de drenagem de 1,67 km/ km².

Os regimes dos rios de Santa Catarina são caracterizados por chuvas bem distribuídas o ano inteiro e garantem uma provisão de água normal durante todo o ano. O comportamento da grande maioria dos rios em conformidade com a distribuição das chuvas é representado através de dois máximos (primavera e no fim do verão) e dois mínimos (registrados no princípio do verão e no outono e algumas vezes adentrando o inverno) características típicas de um regime subtropical.

As características da *vegetação* do Estado de Santa Catarina eram representadas originalmente por regiões fito-ecológicas e áreas de formações pioneiras:

- Floresta Ombrófila Densa (floresta pluvial da costa atlântica);
- Floresta Ombrófila Mista (floresta com pinheiros);
- Floresta Estacional Decidual (floresta latifoliada do Alto Uruguai);
- Savana (campos do planalto meridional);
- Vegetação pioneira de influência flúvio-marinha e marinha (mangue e restinga).

Na Região de Joinville ocorrem a região fitoecológica Floresta Ombrófila Densa (Floresta Atlântica) e uma área de formação pioneira denominada Vegetação Pioneira que são descritas a seguir.

A região fitoecológica Floresta Ombrófila Densa (Floresta Atlântica) está situada na parte leste do Estado de Santa Catarina entre o planalto e o oceano,

constituída na sua maior parte, por árvores perenifoliadas de 20 a 30m de altura, com os brotos foliares sem proteção à seca. Sua área é formada por planícies litorâneas e principalmente por encostas íngremes da serra do Mar e da serra Geral formando vales profundos e estreitos.

O *bio-clima* é caracterizado por precipitações abundantes e regularmente distribuídas durante o ano com um período mais intenso no verão. A umidade relativa é bastante elevada (84 a 86%) nas proximidades da costa diminuindo em direção ao oeste.

Trata-se portanto de um clima azonal tropical úmido, sem período seco e com médias térmicas nunca inferiores a 15°C.

As condições ambientais desta região favorecem ao desenvolvimento de uma mutiplicidade de subformas de vida com grande força vegetativa e alto volume de biomassa, sobretudo tomando-se em consideração o volume de madeira por hectare.

Atualmente grande parte da Floresta Ombrófila Densa encontra-se intensamente explorada pela retirada das madeiras de lei ou completamente devastada para ceder lugar à agricultura ou às pastagens "naturalizadas". O fato de grande parte dessa vegetação se encontrar em terrenos bastantes dissecados e de difícil acesso, existem ainda áreas consideráveis com a cobertura florestal original.

Esse uso intensivo por atividades agrícolas e pastoris leva a uma exaustão do solo e consequente abandono da área utilizada. Denomina-se de vegetação secundária as associações vegetais, compostas por densas aglomerações de ervas, arbustos e árvores de pequeno, médio e grande porte que se sucedem a todo esse processo de utilização-abandono.

A vegetação secundária na área em questão, começa com ervas anuais e termina em floresta, cujo aspecto fisionômico é muito semelhante à floresta original, sendo entretanto sua composição florística bem distinta.

A exploração econômica da Floresta Ombrófila Densa ocorreu com um pequeno número de espécies arbóreas de grande valor econômico tais como a canela-preta; a canela sassafrás; a canela-amarela; a peroba-vermelha; o pau-óleo e outras.

O palmito constitui uma das grandes potencialidades de manejo para o enriquecimento florestal tendo em vista o seu potencial econômico para o mercado interno e externo.

A área de Vegetação Pioneira compreende uma formação vegetal onde predominam espécies herbáceas e arbustivas que são direta ou indiretamente influenciadas pelo mar. Ela é resultado das condições edáficas, tem características bem distintas variando em função de seus diversos habitats que geralmente são agrupados em vasosos, arenosos, rochosos e lagunares.

Nas baías, reentrâncias do mar e desembocaduras dos rios considerados ambientes vasosos, desenvolvem-se a "formação de manguezal" onde predominam espécies arbustivas e pequenas árvores como a siriúba (*Avicennia schaueriana*) dominando fitofisionomicamente; o sapateiro ou mangue-branco (*Laguncularia racemosa*); o mangue-vermelho ou mangue-charuto (*Rhizophora mangle*) e os capins praturás (*Spartina densiflora* e *S. alterniflora*).

No ambiente arenoso encontramos nas praias (ou formação psamófila) as espécies características tais como a salsa-da-praia ou batateira-da-praia (*Ipomea pescaprae*); a acariçoba ou corcel (*Hydrocotyle bonariensis*); a grama-da-praia (*Paspalum vaginatum*); o capotiraguá (*Phyloxerus portulacoides*); o marmeleiro-da-praia (*Dalbergia ecastophyllum*) entre outras.

Nas dunas semifixas dominam em geral a aroeira-vermelha (*Schinus terebinthifolius*) e o pau-de-bugre ou aroeira (*Lithraea brasiliensis*); acompanhados pela capororoca (*Rapanea parvifolia*); a maria-mole (*Guapira opposita*); o guamirim (*Gomidesia palustris*) bem como a vassoura-vermelha ou vassourão (*Dodonaea viscosa*), por vezes bastante frequente, predominantemente em locais onde a vegetação original foi devastada, formando uma vegetação arbustiva alcançando 2 a 3m.

Os terrenos arenosos mais firmes e menos ondulados são ocupados com uma maior frequência pelo guamirins (*Eugenia catharinae*, *E. umbelliflora* e *Gomidesia palustris*) e o cambuí (*Myrcia multiflora*) formando agrupamentos arbustivos geralmente densos.

Nos solos rochosos são frequentes o mangue-de-formida (*Clusia criuva*); a figueira-mata-pau (*Coussapea schottii*); a arumbera (*Opuntia arechavaletae*) e o

mandacaru (*Cereus peruvianus*), além de bromeliáceas rupícolas dos gêneros *Dyckia* e *Aechmea*.

Em relação a *climatologia* a atuação de diversos fatores (radiação solar, latitude, continentalidade, massas de ar, correntes oceânicas) condicionam os elementos climáticos como: temperatura, precipitação, umidade do ar, pressão atmosférica que por sua vez definirão os tipos climáticos.

Para maioria dos parâmetros climáticos, os fatores ambientais mais significativos são as influências oceânicas e topográficas.

No elemento climático temperatura evidencia-se no território catarinense o clima subtropical onde os valores médios anuais definem a mesotermia, apresentando valores que variam de 21,8°C (Itajaí) no litoral, a 13,0°C (São Joaquim) no planalto. O município de Joinville apresenta um valor médio anual de 21,4°C.

Quanto ao parâmetro precipitação temos que o total anual para o Estado de Santa Catarina indica que as isoietas de maiores valores aparecem no oeste, com declínio em direção ao vale do rio do Peixe, contornado respectivamente pelos valores de 2.400mm, 2.000mm, 1.800mm e 1.600mm.

A amplitude pluviométrica no Estado é de 1.154mm, considerando as estações de Xanxerê (2.373mm) no oeste e Araranguá (1.219mm), no litoral. Joinville apresenta uma média anual de precipitação de 2.170,5mm.

No geral a pluviosidade está bem distribuída no território catarinense, devido às atuações do relevo, da Massa de Ar Polar Atlântica e da Massa Tropical Atlântica, que por sua constância, fazem que não ocorra uma estação chuvosa e uma estação seca. Fica definido, pela distribuição durante todo o ano da chuva, o regime tropical.

4.1.2 Situação Ambiental

O município de Joinville, importante pólo industrial de Santa Catarina e responsável por mais de 20% da produção estadual tem uma população de 397.951 habitantes (Fundação IBGE - 1997) dos quais 95% residentes na área urbana.

Com uma população crescendo a uma média de 6,42% ao ano, nos últimos 15 anos, grandes alterações no meio ambiente têm ocorrido, através do aumento

de impermeabilização das novas áreas urbanas, aumentando a velocidade de escoamento superficial. Aliado a esse fato, a remoção da cobertura vegetal origina o carreamento de partículas para os canais de drenagem pluvial ocasionando frequentes cheias no município.

A falta de infra-estrutura urbana adequada e a ineficácia de instrumentos de controle normativo ambiental fizeram com que esse crescimento acelerado causasse grandes degradações ambientais através da ocupação das encostas dos morros, das áreas de mangue e das áreas de fundo de vales.

Paralelamente à ocupação urbana, a industrialização do município também provocou graves alterações ambientais. O lançamento de efluentes industriais nos corpos de água; a extração de madeira; a mineração de seixos, tudo isso tem contribuído para a poluição dos rios e a degradação dos solos.

Reconhecendo a importância dos ecossistemas de mangue como um mantenedor de diversas formas de vida, a administração do município de Joinville estabeleceu diversos mecanismos para conter a destruição desses ambientes que são: caracterização das áreas de mangue; delimitação dessa área; estudo de viabilidade e elaboração de projeto de abertura de canais de contenção da invasão; definição de mecanismos de proteção das áreas de mangue e saneamento das áreas ocupadas onde se caracteriza a inviabilidade da recuperação, acompanhado de uma fiscalização extensiva e elaboração de um projeto de implantação de uma estação ecológica no mangue na região.

O Plano Diretor de Uso do Solo do Município de Joinville, instituído pela Lei nº 1262/73 e depois modificado pela Lei nº 1610/75, determinava como sendo áreas urbanas de preservação permanente todas aquelas situadas acima da cota altimétrica de 40 metros. O artigo 37 da referida Lei, que trata do assunto, foi alterado pela Lei nº 1867/82, ficando aquelas áreas restritas aos morros do Boa Vista e Iririu.

Tendo em vista que a maioria da área urbana não ultrapassa a cota 100, a ocupação urbana tem sido intensa nas pequenas elevações dentro da área urbana. Visando minimizar esse problema a municipalidade desenvolve as seguintes ações:

- levantamento das áreas acima da cota 40;
- cadastramento de proprietários nessas áreas;

- estudos de mecanismos para delimitação física das áreas de preservação permanente;
- estudo de viabilidade de recuperação das áreas degradadas;
- definição de um zoneamento diferenciado;
- definição para aproveitamento de áreas de lazer compatíveis com áreas de preservação permanente.

4.2 A Base de Dados

Duas imagens em formato digital do satélite Landsat (Thematic Mapper) foram utilizadas neste estudo. A primeira na órbita WRS 220.078 em 01 de março de 1990 e a segunda na mesma órbita em 18 de julho de 1994.

As bandas 3, 4 e 5 destas foram utilizadas (3 e 4; 4 e 5) na geração da imagem Distância (variável dependente), resultado da aplicação da Análise do Vetor de Mudança e para produzir imagens compostas de falsa cor para auxílio no registro espacial.

Os planos de informação digitalizados e utilizados foram extraídos de cartas topográficas na escala 1:50.000 - publicado pelo IBGE (Instituto Brasileiro Geografia e Estatística - 1981).

Estes planos de informação compreendem as divisões municipais; as sedes dos municípios; as rodovias, estradas vicinais e ruas; a bacia e as sub-bacias hidrográficas; os rios (hidrologia); as curvas de nível (hipsometria); a geologia; a geomorfologia e as coordenadas geográficas.

O Dados Censitários usados foram providos pelos Anuário Estatístico do Estado de Santa Catarina - compilados pela Secretaria de Estado do Desenvolvimento Econômico, Científico e Tecnológico - Diretoria de Estatística e Geoprocessamento com dados gerados pelo IBGE (Instituto Brasileiro Geografia e Estatística). Este dados foram obtidos pelos censos efetuados em 1991 e 1995.

Os programas de computador utilizados para manipulação e análise dos dados foram:

- PC ARC/INFO® versão 3.5 (ESRI) → Sistema de Informação Geográfica predominantemente vetorial;

- IDRISI for Windows versão 1.01.002 (Clark University©) → Sistema de Informação Geográfica predominantemente raster;
- MATLAB® for Windows versão 4.0 (Copyright© 1984 - 1993 The MathWorks, Inc.) → ambiente computacional matemático com linguagem própria;
- Statistica for Windows Release 5.0 Copyright© StatSoft, Inc. 1984 - 1995 → pacote estatístico para Windows;
- Microsoft Access versão 2.00 Copyright© 1989 - 1994 Microsoft Corporation → Banco de Dados Relacional compatível com o GIS raster.

4.3 Preparação e Processamento dos Dados

4.3.1 Variáveis independentes

A **figura 8** na seção 3.3 apresenta os procedimentos efetuados para obtenção das variáveis *independentes*.

O objetivo final de produção das variáveis *independentes* é a geração dos mapas derivativos ("full surface") que serão amostrados pelo plano vetorial da localização dos pontos de amostragem.

As fontes de dados principais para a geração dessas variáveis são os mapas temáticos e cartográficos, os anuários estatísticos (censos) e os levantamentos de campo.

A partir dessas fontes de dados os planos de informações vetoriais de GIS foram elaborados através de digitalização; conversão de formatos digitais de planos de informações pré-existent e transformações de planos digitais com procedimentos de GIS.

Através de procedimentos de GIS esses planos vetoriais são transformados nos mapas derivativos através de, por exemplo, *interpolações*, *rasterização*, *distância* e *atribuir* (*assign*).

Informação Censitária

Os dados censitários são utilizados para a formação de uma base de dados alfa-numérica no banco de dados relacional para serem utilizadas na formação dos mapas derivativos.

O plano de informação vetorial com as divisões políticas (municípios) da área de estudo é composto de polígonos com um identificador único (**figura 11**). Esse plano é transformado para o formato raster gerando uma imagem final ("full surface") que representa a divisão política da área de interesse. Esta imagem contém polígonos com identificadores que serão usados para se correlacionar com uma tabela de atributos de polígonos no banco de dados relacional.

Em paralelo a esse procedimento, dados censitários de interesse são introduzidos em um banco de dados relacional (**figura 12**) e, através do módulo *atribuir* cada campo de dados (representando uma variável específica) no banco pode ser usado para exibir esse atributo na imagem da divisão municipal correlacionando-se com os identificadores nesse mapa derivativo.

Os dados alfa-numéricos de cada municipalidade adquirem uma dimensão espacial onde os valores de variáveis alfa-numéricas de interesse são distribuídos dentro de seus limites políticos. Esses dados são representados através de pixels com atributos únicos para cada área específica.

As **tabelas 4, 5 e 6** apresentam os dados numéricos utilizados pelo banco de dados relacional para geração das imagens correspondentes à cada variável de interesse. Foram gerados 19 imagens correspondentes às 19 variáveis independentes do banco de dados relacional.

Planos Vetoriais

Os planos de informações vetoriais tais como as rodovias, os rios e as sedes dos municípios foram transformados com a função *distância* em mapas derivativos que contém a distância relativa de cada pixel da característica de interesse (rodovias, rios e sedes dos municípios). Por exemplo, a distância da sede do município, das estradas e dos rios. Essas variáveis foram denominadas respectivamente de: **disvi** (distância das vias); **disau** (distância das sedes dos municípios); **dishd** (distância dos rios)

O plano de informação da hipsometria (curvas de nível) foi transformado de vetorial para raster gerando um conjunto de contornos digitais. Em seguida, através de uma interpolação linear, um Modelo de Elevação Digital (DEM) é gerado contendo a variação de altitude em toda a superfície contínua ao invés de somente isolinhas (**Figura 13**). A denominação dessa variável é **dem**.

4.3.2 Variável dependente

Delimitação da área e correção da Imagem

Das imagens de satélite da órbita WRS 220.078 (2505 colunas X 2240 linhas) dos anos de 1990 e 1994 foi retirado uma janela de amostragem (802 X 587 pixels) com as coordenadas planas 695676 (x1), 719735 (x2), 7099298 (y1), 7116915 (y2) que se enquadrasse dentro dos limites da bacia hidrográfica do Rio Cubatão na Região de Joinville.

A **Figura 14** mostra a região escolhida que tem grande representatividade ambiental pois contém áreas de florestas, de mangue, de agropecuária e urbanizadas.

Essa imagem servirá também como área de amostragem para auxílio nos processos de correção geométrica (registro espacial).

Uma vez que a maioria dos dados vetoriais disponíveis (i.e., hidrografia, estradas e ruas, geomorfologia, hipsometria) eram na circunscrição da bacia procurou-se delimitar uma área da variável dependente que correspondesse à mesma área das variáveis independentes disponíveis. Além disso, o processo de amostragem deve corresponder pixel a pixel para que os dados tenham significância.

Um problema complexo quando se trabalha com imagens de satélite é a questão de sua correção geométrica discutida na seção 2.6.2. Para se fazer a reamostragem e corrigir a imagem em relação às coordenadas geográficas ("rubber sheet transformation") é necessário definir pontos de amostragem, que devem ser praticamente ao nível de pixel, operação trabalhosa e sujeita a erros.

Essa operação de correção geométrica foi feita na imagem amostrada em relação aos planos vetoriais para que a variável dependente tivesse as mesmas

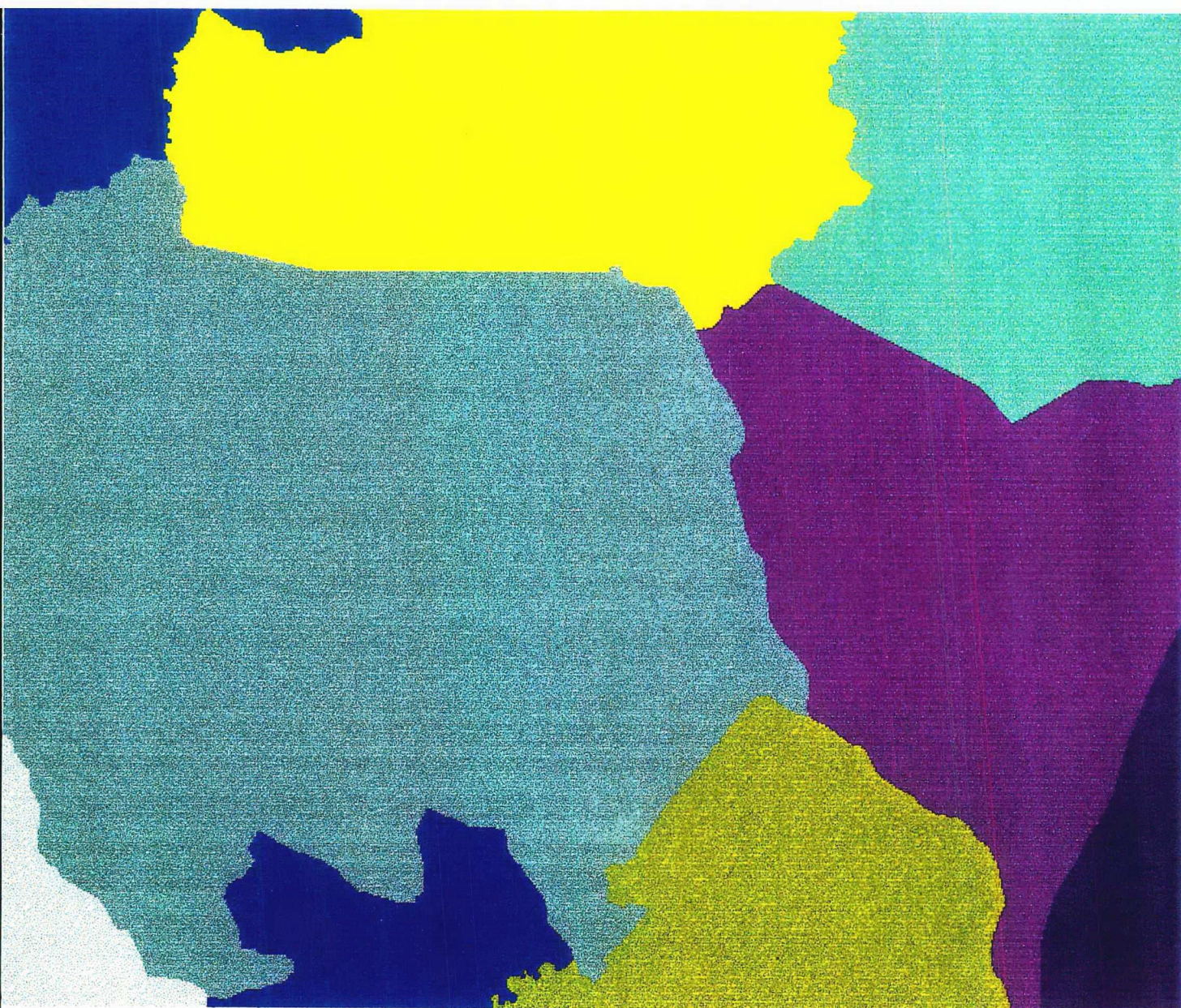


Figura 11: Plano de Informação Vetorial com as Divisões Políticas

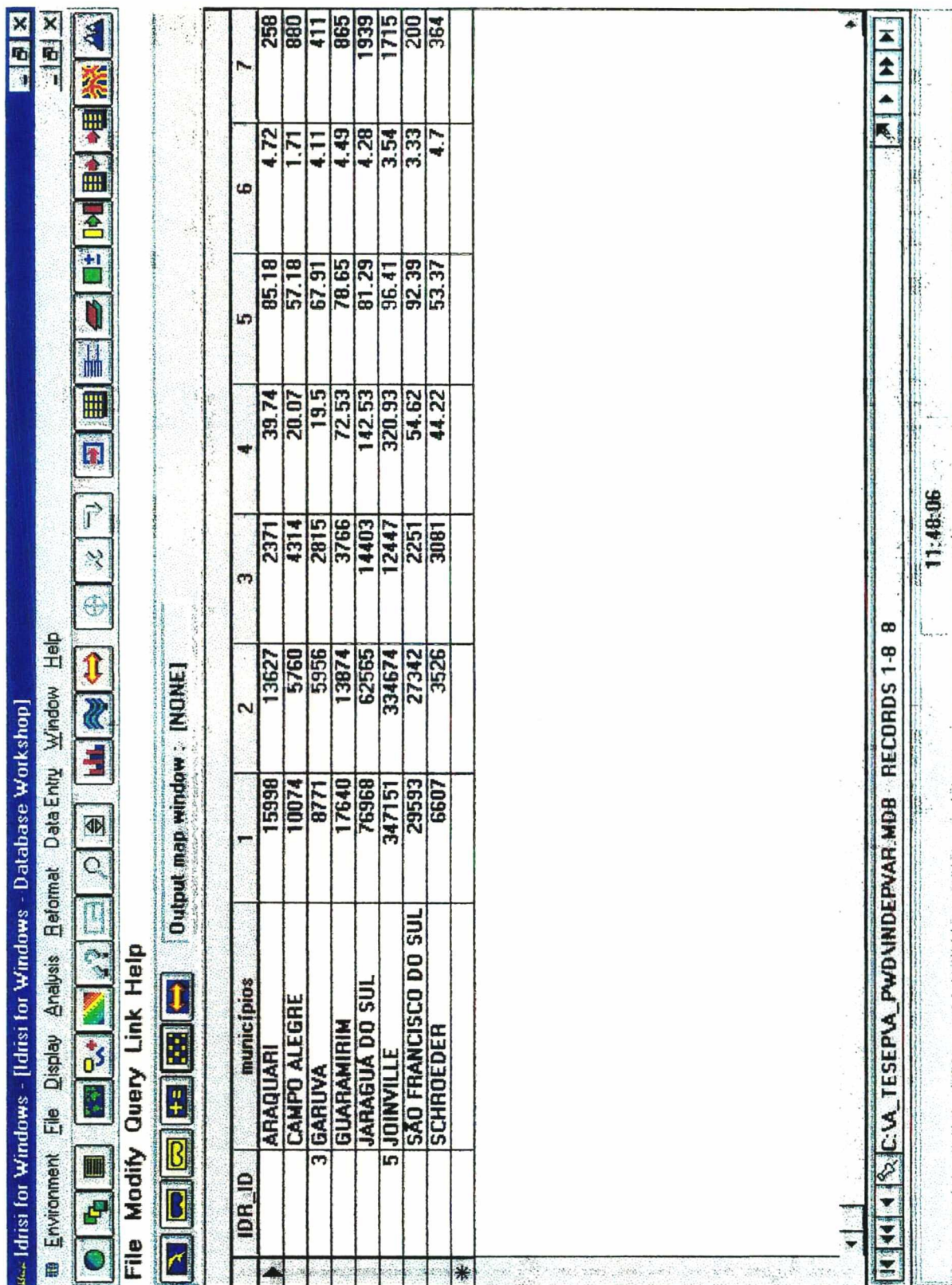


Figura 12 Tabela do Banco de Dados Relacional

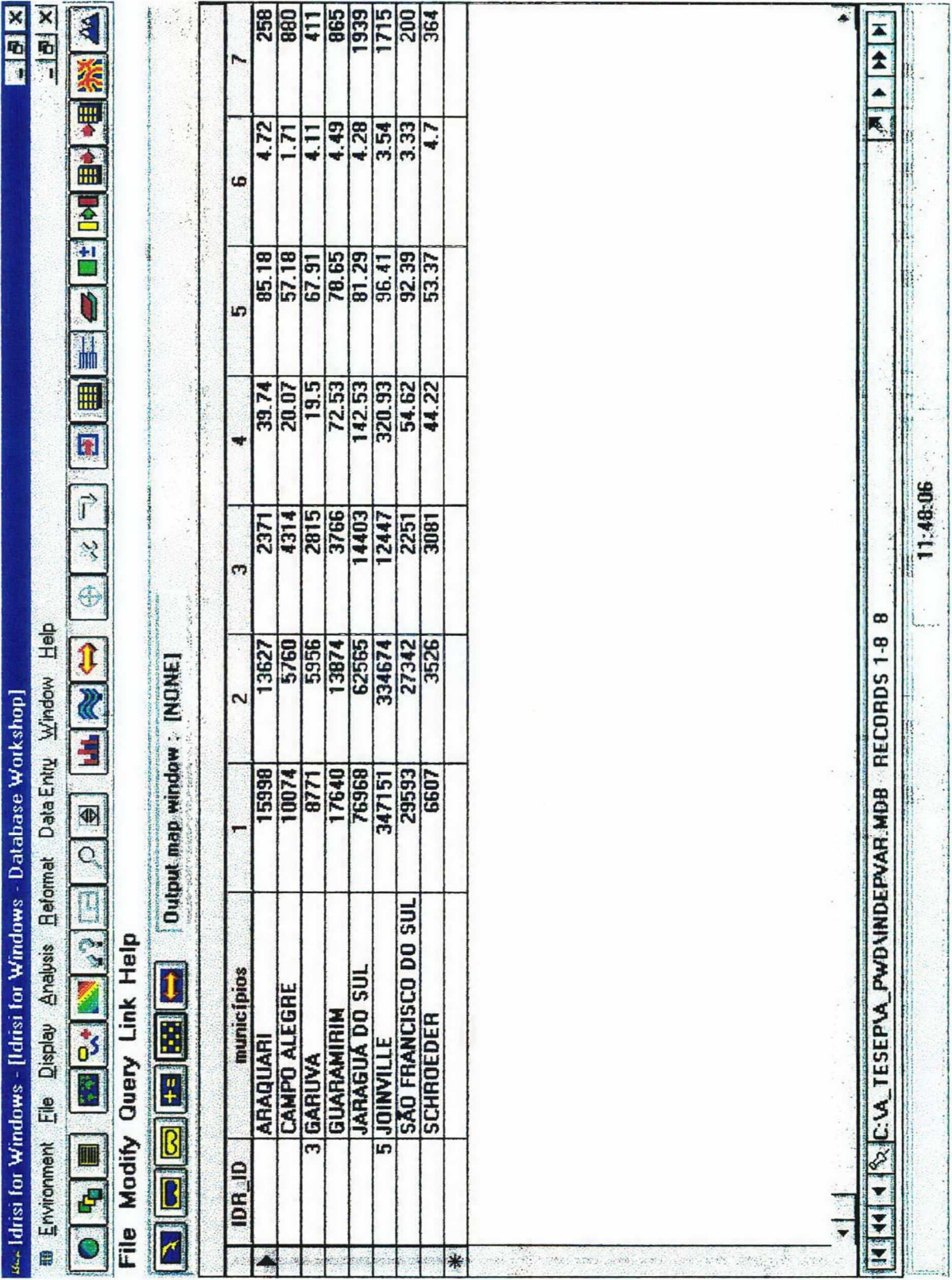


Figura 12 Tabela do Banco de Dados Relacional

| VARIÁVEIS SOCIAIS | população (1) | pop. urbana (2) | pop. rural (3) | densidade demográfica (4) | urbanização (5) | taxa de incremento anual da população (6) |
|----------------------|---------------|-----------------|----------------|---------------------------|-----------------|---|
| | | | | hab/km2 | | |
| ARAQUARI | 15998 | 13627 | 2371 | 39.74 | 85.18 | 4,72 |
| CAMPO ALEGRE | 10074 | 5760 | 4314 | 20.07 | 57.18 | 1,71 |
| GARUVA | 8771 | 5956 | 2815 | 19.50 | 67.91 | 4,11 |
| GUARAMIRIM | 17640 | 13874 | 3766 | 72.53 | 78.65 | 4,49 |
| JARAGUÁ DO SUL | 76968 | 62565 | 14403 | 142.53 | 81.29 | 4,28 |
| JOINVILLE | 347151 | 334674 | 12447 | 320.93 | 96.41 | 3,54 |
| SÃO FRANCISCO DO SUL | 29593 | 27342 | 2251 | 54.62 | 92.39 | 3,33 |
| SCHROEDER | 6607 | 3526 | 3081 | 44.22 | 53.37 | 4,70 |

Tabela 4: Variáveis Sociais

| VARIÁVEIS ECONÔMICAS | proprieda des rurais (7) | proprieda des rurais / área (8) | % ocupado por área agrícola (9) | área de proprieda des <10 ha (10) | área de proprieda des entre 10 e 20ha (11) | área de proprieda des entre 20 e 50ha (12) | área de proprieda des entre 50 e 100ha (13) | área de proprieda des entre 100 e 200ha (14) | área de proprieda des entre 200 a 500ha (15) | área de proprieda des > 500ha (16) |
|----------------------|--------------------------|---------------------------------|---------------------------------|-----------------------------------|--|--|---|--|--|------------------------------------|
| | | | % | % | % | % | % | % | % | % |
| ARAQUARI | 258 | 0.64 | 60.59 | 0.64 | 1.86 | 6.15 | 3.10 | 2.64 | 13.00 | 37.64 |
| CAMPO ALEGRE | 880 | 1.75 | 65.26 | 3.13 | 5.58 | 15.78 | 11.17 | 9.32 | 8.56 | 11.72 |
| GARUVA | 411 | 0.91 | 66.24 | 1.35 | 2.26 | 7.65 | 5.01 | 4.94 | 2.20 | 42.82 |
| GUARAMIRIM | 865 | 3.56 | 77.55 | 5.63 | 15.76 | 24.33 | 9.91 | 3.82 | 4.68 | 13.40 |
| JARAGUÁ DO SUL | 1939 | 3.59 | 56.00 | 6.74 | 18.62 | 21.70 | 2.70 | 1.74 | 0.53 | 3.96 |
| JOINVILLE | 1715 | 1.59 | 37.81 | 2.18 | 5.66 | 10.98 | 3.15 | 2.37 | 3.46 | 10.01 |
| S. FRANC. DO SUL | 200 | 0.37 | 23.52 | 0.36 | 1.04 | 1.72 | 1.98 | 1.78 | 3.07 | 13.56 |
| SCHROEDER | 364 | 2.44 | 45.22 | 3.13 | 14.88 | 20.63 | 3.29 | 1.75 | 1.53 | 0.00 |

Tabela 5: Variáveis Econômicas

| VARIÁVEIS BIO-FÍSICAS | distância da capital | área | altitude | temperatura anual média (17) | precipitação média anual (18) | umidade relativa anual média (19) |
|-----------------------|----------------------|--------|----------|------------------------------|-------------------------------|-----------------------------------|
| | km | km2 | m | C | mm | % |
| ARAQUARI | 172 | 402.6 | 9 | 20.23 | 168.15 | 86.08 |
| CAMPO ALEGRE | 253 | 502 | 870 | 20.23 | 168.15 | 86.08 |
| GARUVA | 226 | 449.7 | 25 | 20.23 | 168.15 | 86.08 |
| GUARAMIRIM | 199 | 243.2 | 30 | 20.23 | 168.15 | 86.08 |
| JARAGUÁ DO SUL | 209 | 540 | 29 | 20.23 | 168.15 | 86.08 |
| JOINVILLE | 185 | 1081.7 | 3 | 20.23 | 168.15 | 86.08 |
| SÃO FRANCISCO DO SUL | 194 | 541.8 | 9 | 20.23 | 168.15 | 86.08 |
| SCHROEDER | 213 | 149.4 | 38 | 20.23 | 168.15 | 86.08 |

Tabela 6: Variáveis Bio-Físicas

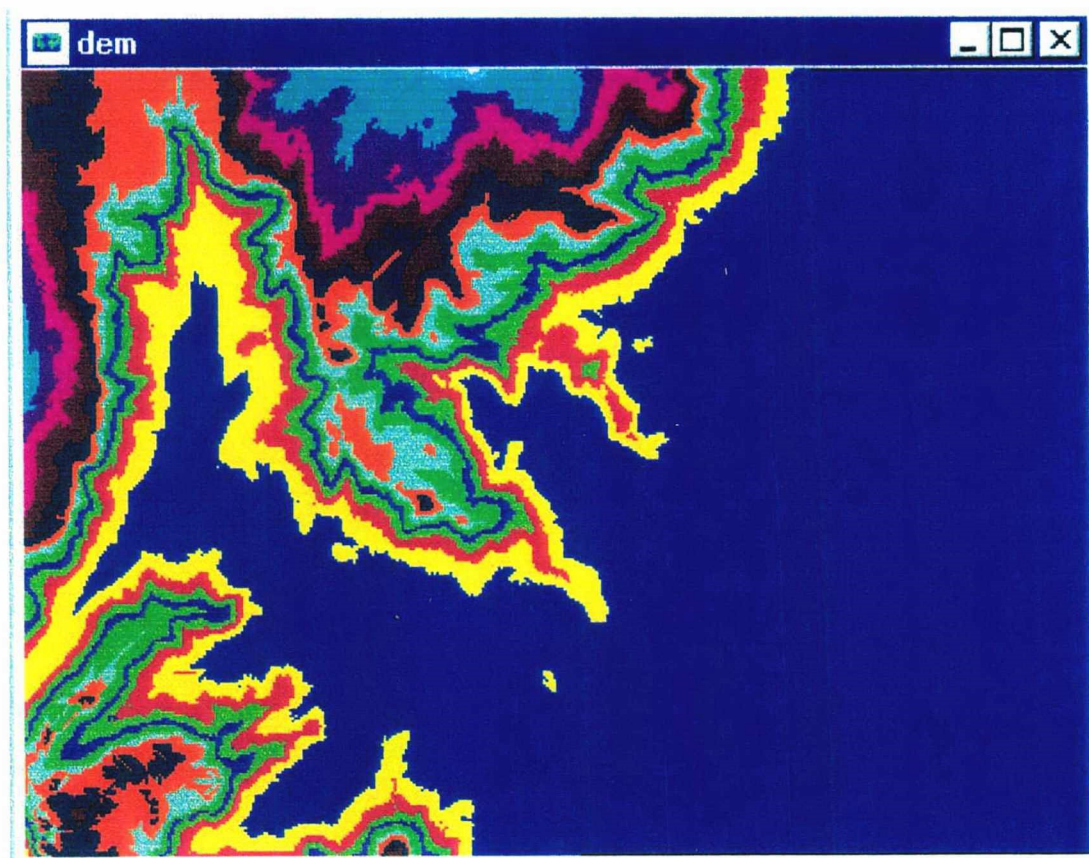


Figura 13: Modelo de Elevação Digital (DEM) gerado por interpolação a partir de isolinhas hipsométricas

coordenadas geográficas das variáveis independentes. O erro RMS ("root-mean-square") dos 35 pontos de controle foi de 0,83 omitindo-se os pontos com residual mais alto para se atingir esse RMS. O RMS das imagens raster deve ser menos da metade da resolução da imagem analisada (Eastman, 1995) que no presente caso é de 30 X 30 m.

Após o registro espacial, da imagem da área em análise, com os planos de informações vetoriais, a imagem de 1994 foi registrada espacialmente com a imagem de 1990 (que já tinha sido registrada com o plano vetorial - registro imagem/ vector) para utilização na Análise do Vetor de Mudança (vide seção 3.3.2).

Como resultado final da utilização dessa técnica de Análise do Vetor de Mudança obtém-se a imagem final (distance.img) que indica a magnitude da mudança na paisagem em estudo (**figura 15**). A **figura 16** mostra o histograma da imagem distance.img onde quanto maior o valor do atributo do pixel (0-255) maior o grau de magnitude de mudança.

4.3.3 Amostragem e Arquivos de Valor

Uma vez que todas as variáveis de interesse já tinham sido geradas como imagens raster de superfície contínua o procedimento subsequente foi a produção de arquivos vetores de localização dos pontos para posterior geração dos arquivos de valor que migram para a RNA.

Esse arquivo vetor foi elaborado selecionando-se um esquema de amostragem estratificado (ao invés da possibilidade do esquema aleatório ou sistemático). Esse esquema compreende:

- inicialmente através da função Sample no GIS gera-se um plano vetorial com arquivos de localização de pontos (arquivos com 50, 100, 200 e 500 pontos de amostragem;
- em seguida, no módulo Initial cria-se uma imagem (sempre com as mesmas coordenadas da área de interesse) raster que servirá para extração dos pontos de amostragem de cada imagem de interesse;

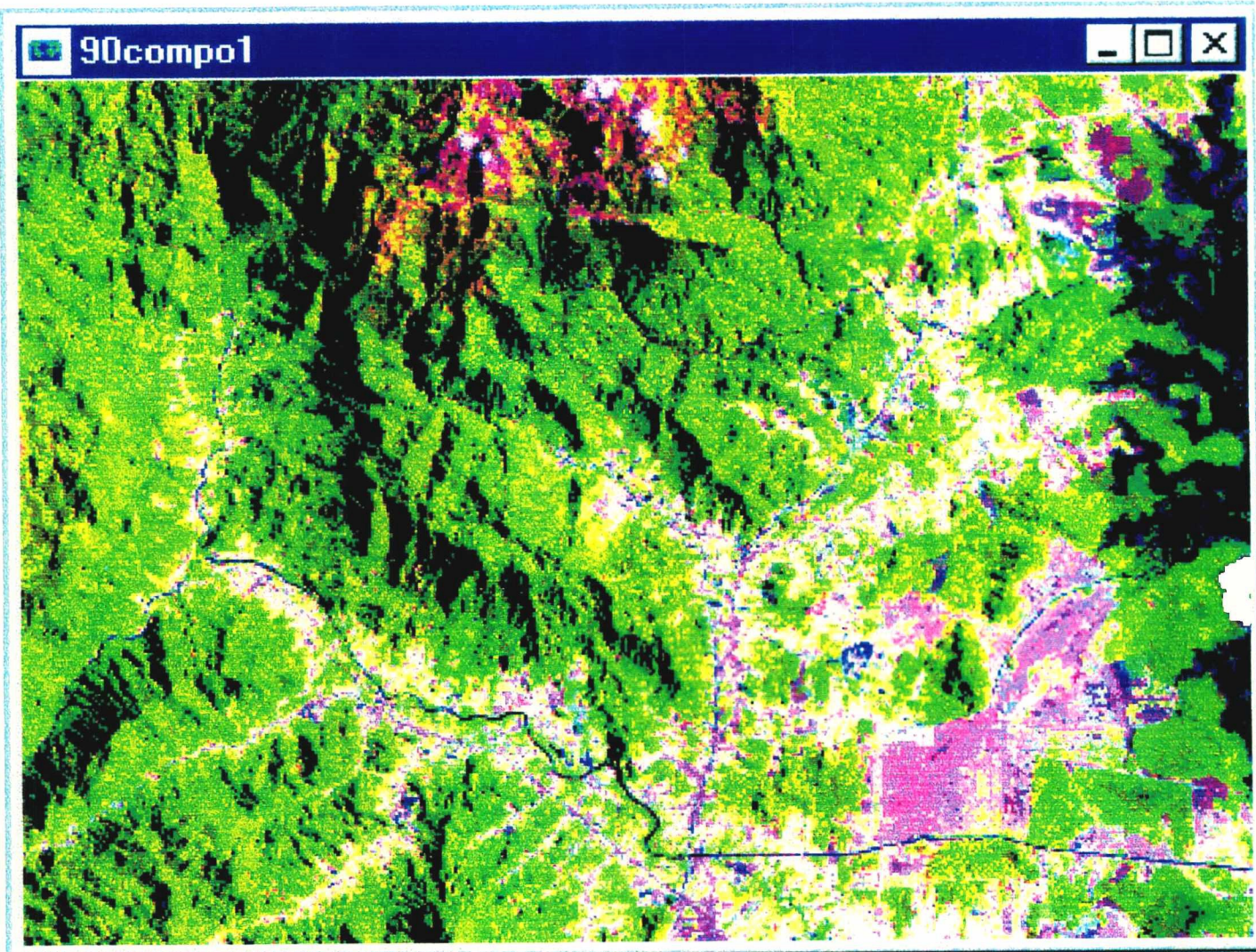


Figura 14: Área da Região do Rio Cubatão - Joinville/SC

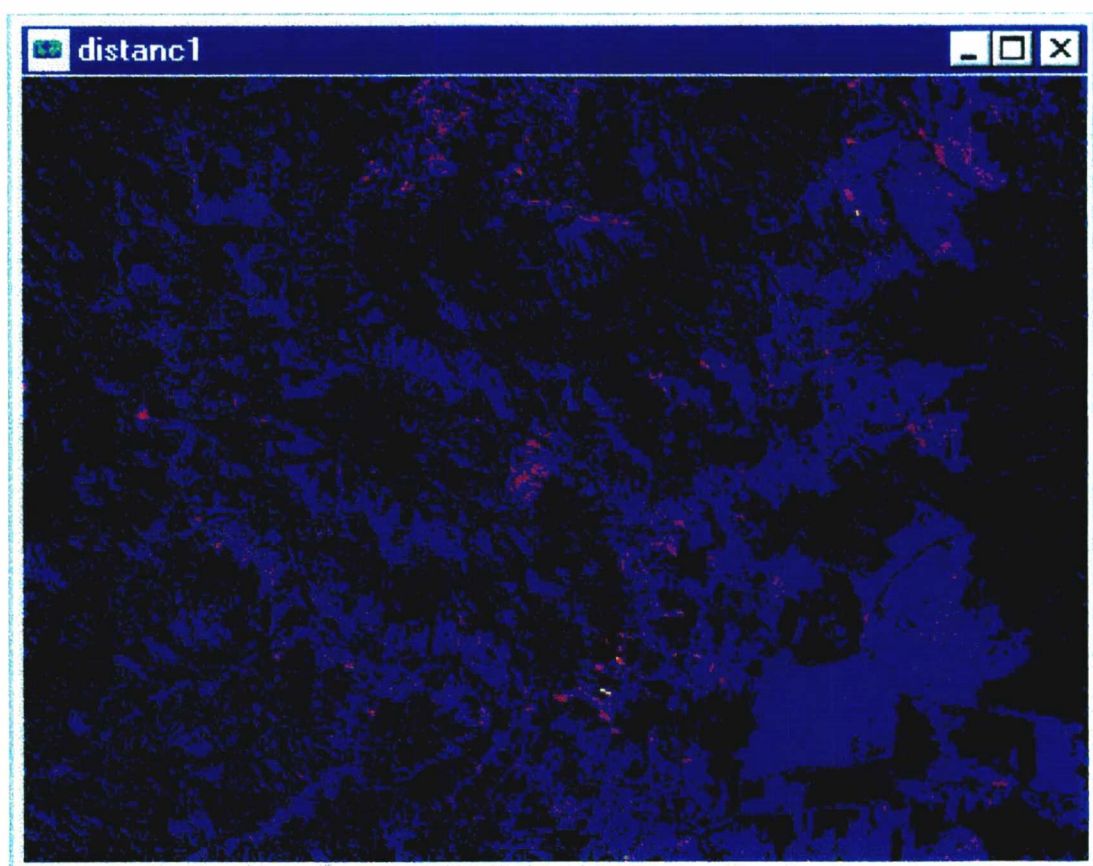


Figura 15: Imagem resultante (distance.img) da Análise do Vetor de Mudança ilustrando a magnitude da mudança

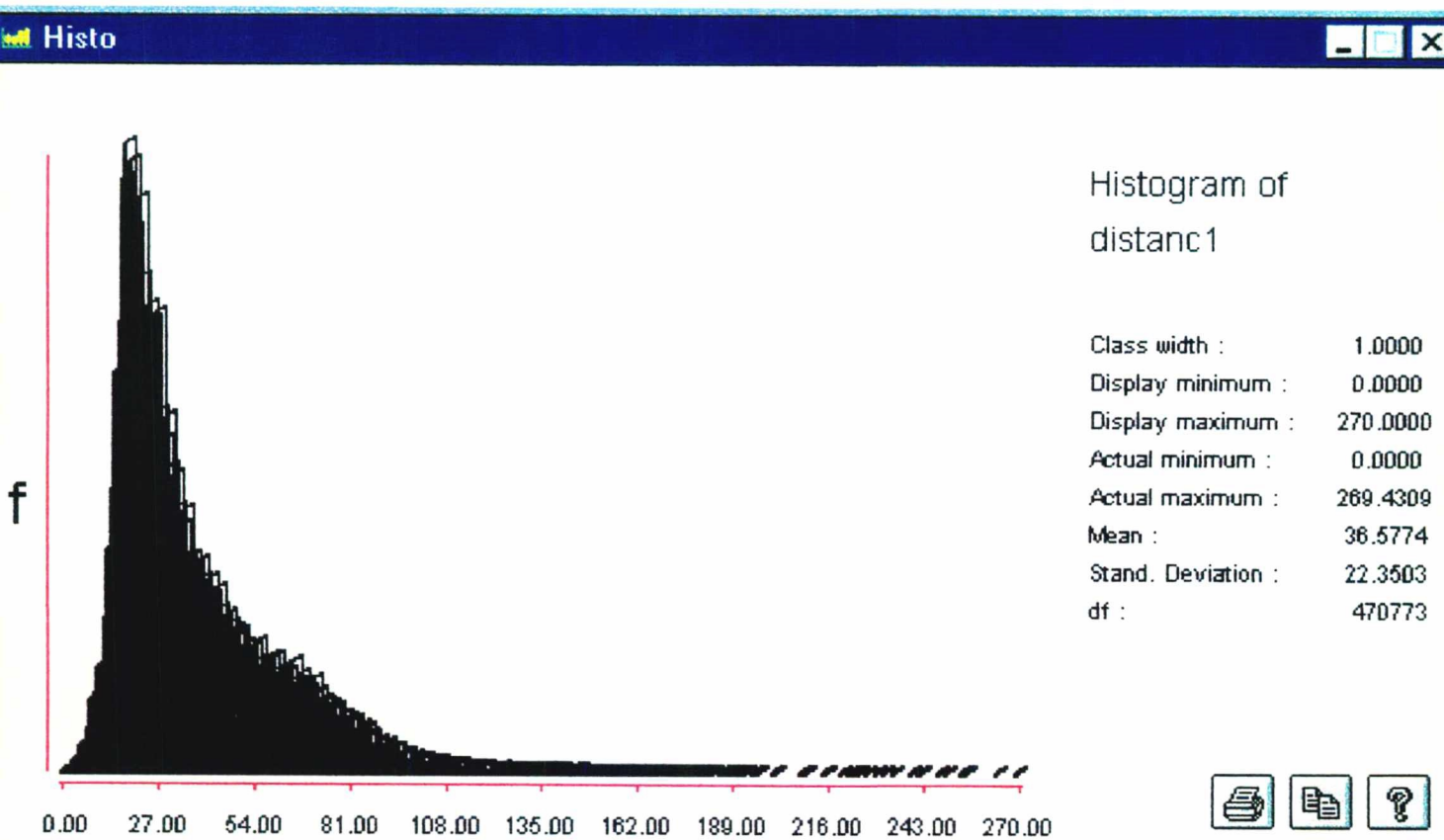


Figura 16: Histograma da imagem distance.img

- no módulo Pointras, que converte dados vetoriais de pontos para uma representação raster, essa nova imagem raster representa os pontos na forma de pixels;

- finalmente, na função Extract utilizando-se da imagem raster de pontos extrai-se os valores amostrais das imagens-variáveis de interesse, gerando a partir daí os arquivos de valores que migrarão para a Rede Neural.

Resumindo, inicialmente, um plano vector de pontos de amostragem é gerado, em seguida transformado em uma imagem de pontos de amostragem para que se possa extrair da imagem da variável de interesse os valores referentes às localizações dos pontos.

A **figura 17** apresenta alguns pontos vetoriais de amostragem sobre uma imagem de variável aumentada várias vezes. É importante salientar como a extração do valor da variável é feita ao nível do pixel claramente distinguível na **figura 17**.

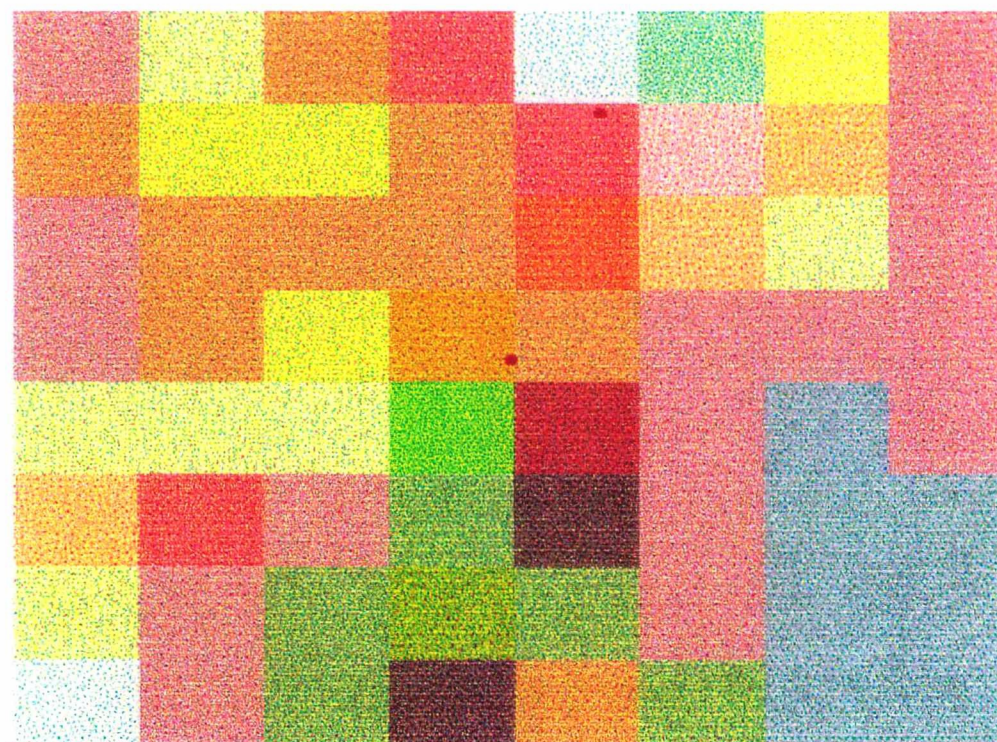


Figura 17: Pontos vetoriais de amostragem sobre uma imagem de variável de interesse

O resultado final são os arquivos de valores em formato ASCII, delimitados por espaços de duas colunas. Esses arquivos podem migrar para o ambiente matemático onde um programa de RNA foi elaborado.

4.3.4 Aplicativo para processamento na RNA

A preparação e o processamento dos dados obtidos do Sensoriamento Remoto e do Sistema de Informações Geográficas migram para um ambiente de computação técnica para análise computacional e visualização de alta-performance (MATLAB® for Windows).

Esse ambiente computacional também contém uma série de módulos ('toolboxes') de coleção de funções (M-files) que potencializam o ambiente matemático para resolver determinadas classes de problemas. O módulo de Redes Neurais foi utilizado para esse trabalho.

Dentro desse ambiente matemático uma macro-linguagem interna permite o desenvolvimento de aplicações sem a necessidade de utilização de linguagens de baixo nível (i.e., Fortran).

Dentro desse contexto, uma aplicação foi desenvolvida na linguagem interna para o processamento dos arquivos de valor ou de imagens através de uma Rede Neural Artificial no módulo específico.

Aplicação

As figuras que ilustram o texto a seguir compreendem janelas ('windows') que vão abrindo conforme determinadas tarefas são requisitadas e elaboradas dentro da aplicação.

A **figura 18** representa a tela inicial com as funções principais do aplicativo.

Neste ponto o usuário pode iniciar a preparação dos dados ("data preparation"); padronizar os dados dentro de uma determinada proporção ("scaling"); iniciar ou reiniciar o treinamento da rede ("training"); quando a rede já estiver treinada pode testá-la com dados novos ("testing"); uma janela com uma série de funções auxiliares dentro do processamento ("utilities"); e

finalmente uma última janela que permite retornar à janela anterior quando esse procedimento for necessário.

A **figura 19** apresenta a tela de preparação dos dados, onde existe a opção de imagem ou arquivo de valor. O texto embaixo representa sequências de procedimentos que aparecem na janela de comando do ambiente matemático computacional.

Neste ponto dá-se a preparação dos vetores de entrada ("input") e dos vetores de saída ("target"). Neste momento a arquitetura da rede define-se a respeito dos neurônios de entrada (número total de entradas desejadas) e dos neurônios de saída (número total de saídas desejadas). A camada interna de neurônios ("hidden layer"), como veremos mais adiante, pode ser alterada em função dos resultados obtidos.

Na segunda janela da mesma figura existe a possibilidade de escalonar os dados que serão processados pela rede. No exemplo que aparece da janela de comando (linhas abaixo da segunda janela) esse escalonamento foi feito entre 0,1 e 0,9.

A **figura 20** apresenta as telas subsequentes à função de treinamento ("training"). A primeira janela possibilita processar a rede pela primeira vez e reprocessá-la visando alterar alguns parâmetros e aumentar a sua eficiência.

A segunda janela contém funções que permitem manter os parâmetros por "default" (indicados no texto da janela de comando abaixo) ou a possibilidade de modificação de alguns parâmetros da rede tais como: número de épocas (ciclos), o erro desejado, taxa de aprendizado, momento, o tipo de função de ativação do neurônio.

A **figura 21** contém as funções auxiliares dentro do processamento dos dados.

É possível gravar ("saving") os dados de treinamento e os de teste da rede. Da mesma maneira também é possível reescalonar esses dados para uma nova verificação dentro da arquitetura da rede.

A segunda tela dessa figura permite ao usuário gravar os dados de entrada, de saída ou ambos, tanto para os dados de treinamento como para os dados de teste da rede.

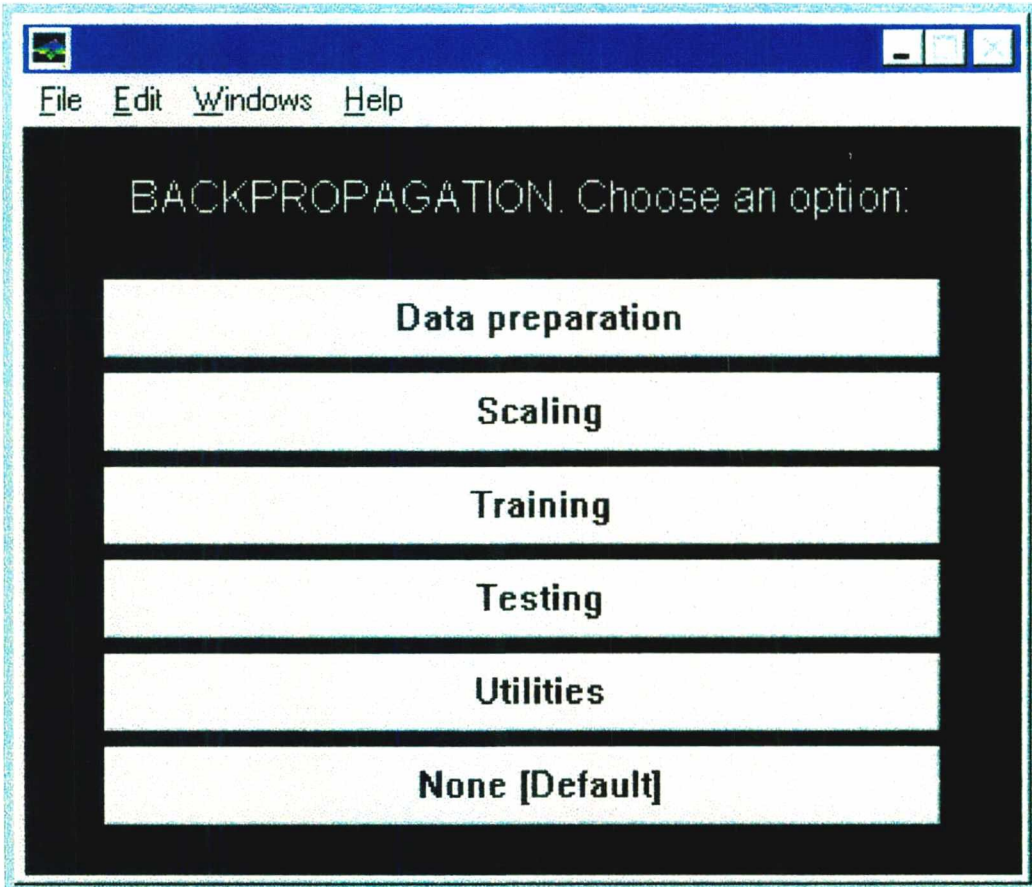
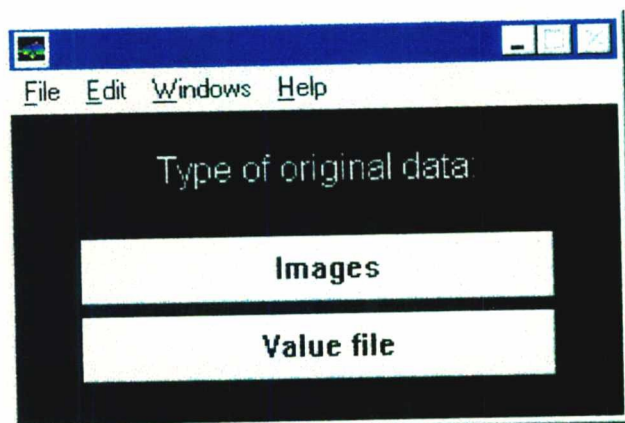


Figura 18: Tela inicial com as funções principais

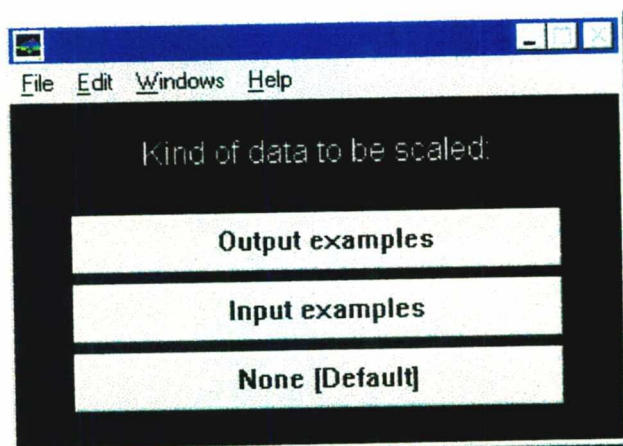


Backpropagation Training

Data preparation

FILE TRANSFORMATION

Input the image name: (into quotes)



Data Scaling

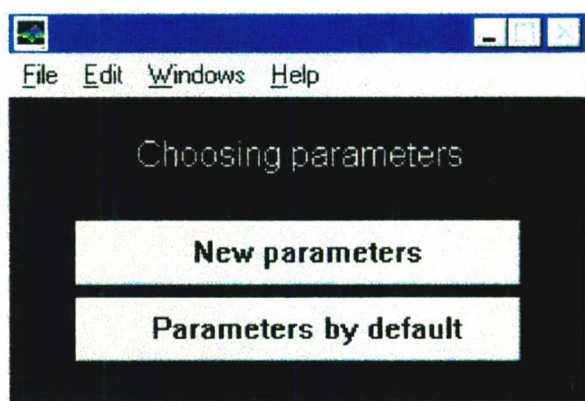
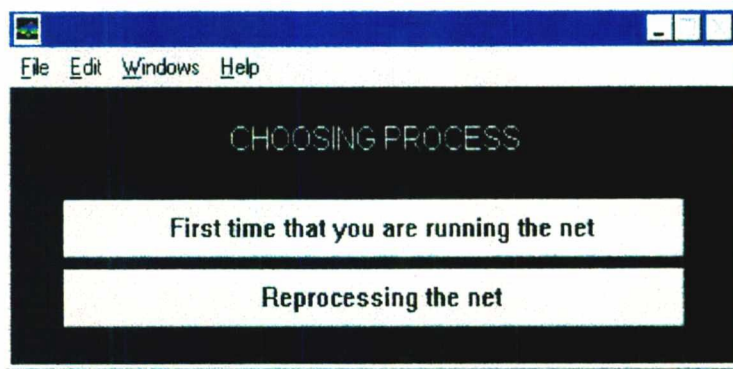
Do you want to scale the values (0 yes / 1 no): 0

Scaling between a range...

Minimum value in the scale: 0.1

Maximum value in the scale: 0.9

Figura 19: Tela com as funções de preparação dos dados



Training

Initialization. Wait.....

Loading the weights.....

PARAMETERS BY DEFAULT ARE:

Display frequency: 10 , Maximum number of epochs: 200 , Error goal: 0.003000 Learning rate: 0.010000 , Learning increment: 1.050000 , Momentum: 0.900000
Act. function 1st. layer: logsig , Act. function 2nd. layer: logsig

The number of hidden neurons by default is: 2

Figura 20: Processamento da rede e modificação de parâmetros

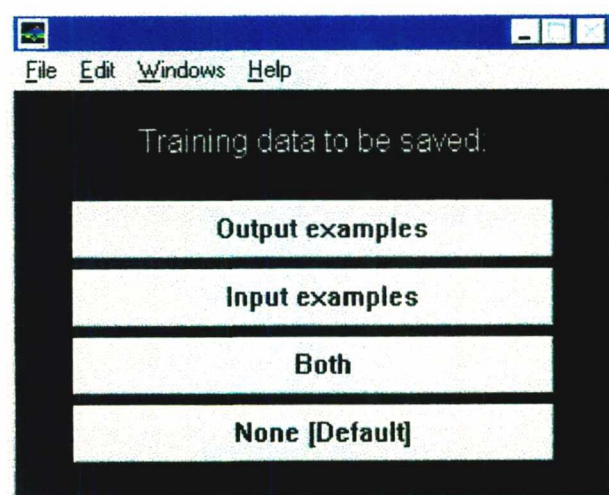
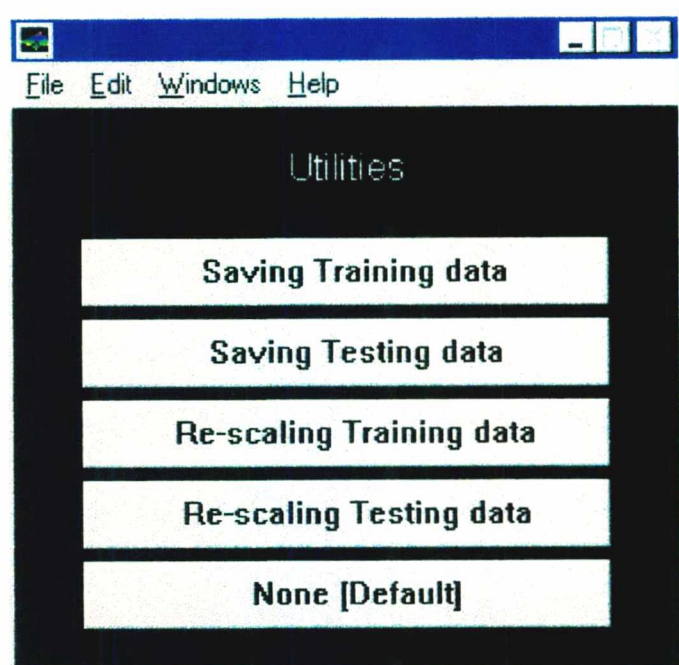


Figura 21: Funções auxiliares dentro do processamento dos dados

Testing a trained network

Testing the network. Wait.....

| | | |
|-------------------|------------------|------------------|
| Desired Output | Actual Output | Squared Error |
|-------------------|------------------|------------------|

End Testing.....

Saving Results (Target output, Actual output, Sum Squared Error)

Figura 22: Resultados da utilização da função de teste da rede

A última figura (**figura 22**) resulta quando se utiliza da função de teste (“testing”) da primeira janela da aplicação. Uma vez que os dados para teste já estejam preparados eles podem ser processados por essa função e o resultado aparece na janela de comando do ambiente computacional.

Com a rede já treinada até um certo ponto, essa função compara a saída desejada com a saída efetiva que a rede produz e calcula o erro residual. Em seguida, esses resultados são automaticamente gravados.

4.4 Análise dos Dados

Como procedimento inicial para treinamento da RNA, a identificação do subconjunto de variáveis (vide seção 4.4.1 para a relação de variáveis) que validarão o modelo tem importância crucial, tanto do ponto de vista do entendimento das possíveis relações de causa-efeito de mudança na paisagem assim como no estabelecimento da arquitetura definitiva da rede.

A definição do número de variáveis determina parte da arquitetura da rede especificando o número de neurônios de entrada e o(s) de saída. Após essa decisão, a única interferência possível na arquitetura da rede é a modificação do número de neurônios da camada interna.

O processo de treinamento da rede é basicamente empírico, com algumas regras intuitivas que podem eliminar passos desnecessários durante o processo.

O tamanho da amostra foi o passo inicial para o treinamento da RNA, após, a definição dos parâmetros que desenvolvessem uma rede mais eficiente com o menor número de neurônios e ciclos possíveis e que tivesse uma convergência para o menor SSE dentro do espaço de erros presente para aquela amostra.

Simulações foram feitas com arquivos de valor (ASCII) contendo 50, 100, 200 e 500 pares de vetores. O tamanho de amostra com 50 vetores alcançou o menor SSE em relação aos outros tamanhos.

Conjuntamente com essa variação analisou-se a influência que o número de neurônios da camada interna teria sobre a eficiência da rede.

Com aquela estrutura de entradas e saídas determinadas (quatro variáveis de entrada e uma de saída), a rede estabilizou-se melhor com quatro neurônios na camada interna (**figura 23**).

Em relação ao número de épocas, quando se comparava o tamanho da amostra procurou-se seguir a orientação de Haykin (1994), para quem o número de épocas é escolhido de maneira que o número total de padrões de treinamento (tamanho da amostra * número de épocas) usados em cada simulação seja constante. Após fixar esses valores para as primeiras

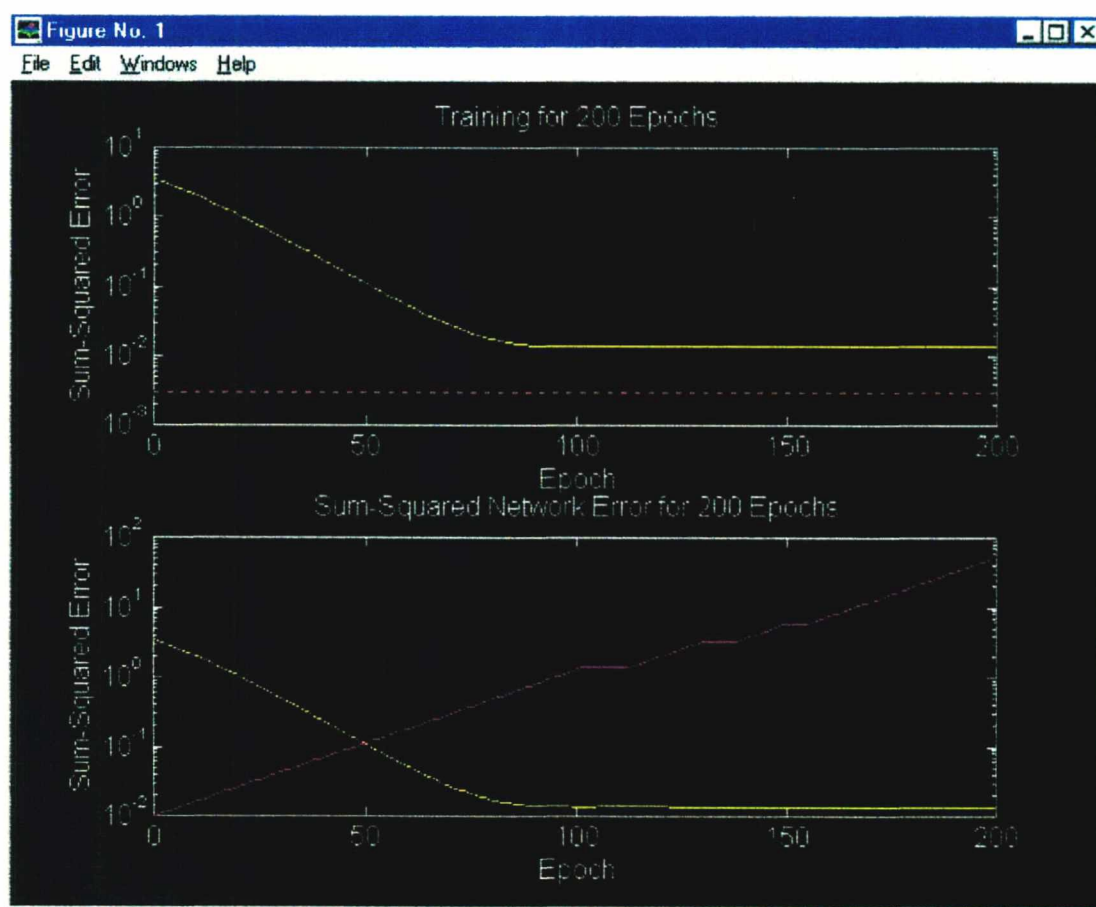


Figura 23: Redução da SSE (amarelo), aumento da taxa de aprendizagem (magenta)

simulações obteve-se o melhor SSE com um número total de 5000 épocas, para um tamanho de amostra com 50 pares de vetores coluna e quatro neurônios na camada interna.

Os parâmetros utilizados por default são: frequência de apresentação (10); número máximo de épocas (200); meta de erro (0,003); taxa de aprendi-

zagem (0,01); incremento da aprendizagem (1,05); momento (0,9); função de ativação na primeira camada (logsig) e função de ativação na segunda camada (logsig).

Os parâmetros que não foram alterados para as várias simulações elaboradas foram a meta de erro e as funções de ativação.

Os melhores resultados de simulação foram alcançados com taxas de aprendizagem menores (0,01) e alterações no valor do momento não ocasionaram alterações perceptíveis na convergência da rede.

Alterações no incremento da aprendizagem também não indicaram qualquer significância que pudesse ser aferida.

Foram elaboradas algumas simulações de teste com dados não vistos pela rede. A capacidade de generalização da rede pode ser melhorada tendo em vista o SSE alcançado até agora, a arquitetura de rede elaborada e a maneira como as variáveis foram compostas para montar o esquema de procura de associação entre os dados (“mapping”).

Novas simulações devem ser elaboradas considerando outros possíveis relacionamentos e conjuntos de variáveis buscando entender os inter-relacionamentos de causa-efeito que compõem e formam o padrão de mudança da paisagem.

É importante considerar também que as variáveis independentes escolhidas têm um grau de correlação baixo com a variável independente como indicam a **tabela 7** e os **anexos 1, 2, 3, e 4**. Entretanto, esse fato não impediu que a rede pudesse convergir, validando o desenvolvimento teórico de que ela pode associar praticamente qualquer relação que exista nos dados independente de sua distribuição estatística, de erros durante o processo de amostragem e outros problemas geralmente encontrados em análises quantitativas.

Na seção seguinte, considerações teóricas e resultados obtidos empiricamente são discutidas à luz de conceitos estabelecidos para as RNA.

4.5 Considerações Finais

| | Var1 | var2 | var3 | var4 | var5 |
|------|-----------|-----------|-----------|----------|-----------|
| var1 | 1 | -0,257293 | -0,347025 | 0,255063 | -0,282523 |
| var2 | -0,257293 | 1 | | | |
| var3 | -0,347025 | | 1 | | |
| var4 | 0,255063 | | | 1 | |
| var5 | -0,282523 | | | | 1 |

Tabela 7: Coeficiente de Correlação entre as variáveis

Haykin (1994) argumenta que a tarefa de aprendizagem imposta à rede neural é que determina grande parte dos procedimentos de aprendizagem. Ele elenca uma série de tarefas de aprendizagem que torna as RNA utilizáveis de uma forma ou de outra.

A primeira é a *aproximação* de uma função não linear entre um conjunto de exemplos pares de vetores de entrada e saída. Esse é um problema candidato ao aprendizado supervisionado e, até se considera que o aprendizado supervisionado é tipicamente um problema de aproximação.

Na tarefa de *associação* de padrões, duas formas são distinguidas. A autoassociação, aprendizado não supervisionado, onde a rede armazena um conjunto de padrões e sua tarefa é recuperar esses padrões quando é utilizada. E a heteroassociação, aprendizado supervisionado, onde o conjunto de pares de vetores de entrada e saída são selecionados arbitrariamente.

A terceira tarefa de aprendizagem, *classificação de padrões*, compreende o processo de classificação de um número de classes presente no conjunto de dados. A rede é treinada com um conjunto de vetores de entradas com os seus respectivos vetores de saída, estes últimos indicando as classes respectivas para cada vetor de entrada. É um processo de aprendizado supervisionado que elabora (mapeia) relações não-lineares e não-paramétricas.

A *classificação de padrões* também pode ser desenvolvida pelo aprendizado não supervisionado. Quando não se conhece as classes que se

pretende classificar esse tipo de treinamento pode fazer o agrupamento (“clustering”) antes de entrar na classificação propriamente dita.

Na tarefa de *previsão*, os vetores de entrada-saída são amostrados de séries históricas e a rede, depois de treinada, deve ter capacidade de prever as saídas para novos valores de entrada apresentados à rede. O processo de treinamento é correção de erro não supervisionado.

O *controle* de processo é mais uma das tarefas que a RNA pode ser induzida a perfazer. As saídas da rede são ações e as entradas são determinadas pelo tipo de controle que se deseja fazer. O tipo de aprendizado aqui é não supervisionado.

A última tarefa sistematizada por Haykin (1994) é um tipo de *filtragem espacial* e o objetivo de aprendizagem é localizar um sinal alvo dentro de um ambiente com interferência adicionada (radares e sonares).

O autor conclui, ao final da sistematização das tarefas de aprendizagem, que as RNA são universais como sistemas de processamento de informações. De uma maneira genérica ele considera que todas essas tarefas são exemplos de mapeamentos elaborados, desde que o problema imposto tenha solução, pelas informações contidas na base de dados.

Durante o processo de aprendizagem supervisionado, os pesos sinápticos são computados durante as simulações de treinamento com um conjunto de dados estacionários e o mais representativo possível do universo da população que se analisa. Quando o treinamento cessa, a rede deve poder capturar o relacionamento estatístico dos dados através da fixação dos pesos, funcionando como uma espécie de memória daquele universo de dados.

Finalmente, Haykin (1994) argumenta que o processo de aprendizado de uma RNA é estocástico por natureza. Ele enfatiza que uma RNA “é meramente uma forma na qual conhecimento empírico sobre um fenômeno físico de interesse pode ser codificado”. Estatisticamente uma RNA é um dispositivo que implementa um modelo matemático regressivo que usa x para explicar ou prever y .

Treinamento

O treinamento inicia-se procurando estabelecer o número mínimo de neurônios da camada interna que possa alcançar a meta de erro estabelecida.

Os parâmetros de taxa de aprendizagem e momento são mantidos constantes neste ponto. Para cada simulação todo o conjunto de treinamento é apresentado à rede e cada época corresponde a uma iteração de todo o conjunto da amostra.

Quando se inicia a rede, os pesos são delimitados aleatoriamente. Da plotagem da função de erro (J) contra os pesos sinápticos obtém-se a superfície do erro para aquele problema. Quando o problema que se aborda é não linear a superfície do erro tem um ou vários mínimos globais assim como mínimos locais (Haykin, 1994).

Basicamente o objetivo do treinamento é começar o processo de um ponto arbitrário da superfície de erro e mover-se na direção do mínimo global.

Algumas vezes, geralmente, quando o problema é mal formulado, a rede pode ficar presa em um mínimo local e nunca conseguir sair dali, não chegando à solução final (mínimo global).

Haykin (1994) indica que um erro “baixo” alcançado pela rede não implica em uma boa generalização da rede. Por generalização entende-se a capacidade da rede em utilizar dados nunca vistos, ou seja dados que não foram utilizados no treinamento, e obter respostas corretas ou próximas do erro mínimo alcançado pela rede. Se a rede tem uma capacidade “alta” de generalização, mesmo com entradas ligeiramente diferentes, ela produzirá um mapeamento correto das entradas-saídas.

Parametrização

Conforme já discutido, o tamanho do conjunto de treinamento; a quantidade de neurônios na camada interna; a taxa de aprendizado; o número de épocas; são fatores que podem ser alterados empiricamente

durante as simulações para se chegar ao objetivo maior de treinamento, que é alcançar o erro mínimo global (quando isso é possível) e ao mesmo tempo assegurando que a rede tem uma boa capacidade de generalização (teste).

Entretanto, como se depreende da literatura, não existem procedimentos único para treinamento de redes. Os enfoques e os procedimentos de treinamento são diversos em relação ao estabelecimento de metas de erro, número de épocas, número de neurônios mínimo ou máximo dentro da camada interna.

Alguns autores trabalham na faixa de 5.000 a 10.000 épocas enquanto que outros em faixas de 150.000 a 200.000 épocas.

Dentre as críticas às RNA, o argumento de que elas sejam como uma caixa preta aonde se coloca algumas entradas e de alguma maneira obtém-se algumas saídas tem certa sustentação. A equação final é desconhecida e o que se obtém é uma matriz de pesos treinada associada a determinada arquitetura de rede e certo número de parâmetros fixos.

Às vezes, mesmo reiniciando o treinamento com os padrões iniciais idênticos de entrada a rede pode obter um resultado final diverso, uma vez que a matriz de pesos gerada aleatoriamente no início do treinamento não tem relação com a outra matriz da primeira simulação.

Da mesma maneira, alguns parâmetros parecem em algumas simulações indicar determinados caminhos convergentes para o treinamento. Entretanto quando se aumenta exponencialmente o número de épocas a taxa de aprendizado "default" opera melhor que taxas de aprendizado menores que tinham sido utilizadas em outras simulações com números de épocas menores.

Em contraposição à essas críticas, o argumento de que as RNA são extremamente consumidoras de tempo e máquina não parece tão plausível com o preço (em US\$) dos equipamentos digitais caindo paulatinamente. A capacidade de máquinas instaladas parece ser extremamente superior à capacidade de elaboração de problemas que possam ser resolvidos pelas RNA.

Quando se fala em computação paralela para resolver esses tipos de problemas subestima-se o potencial instalado em arquiteturas digitais instaladas.

5. Conclusões e Sugestões para Futuros Trabalhos

5.1 Conclusões

Flamm, R.O. e Turner, M.G. (1992) postulam duas considerações básicas investigando a influência que o uso do solo tem na modificação do padrão da paisagem.

Eles indagam se é possível desenvolver alguma forma de representação das informações derivadas de áreas de conhecimento sócio-econômicas e biológicas. A outra questão levantada é se poderia haver uma solução generalizada para os problemas de manejo e conservação da paisagem.

Representação digital

A primeira questão é mais factível e foi abordada durante todo o decorrer do presente trabalho.

A nível digital o unificador da estrutura dos dados é o pixel e, num aspecto mais geral a imagem raster composta de pixels. Aparentemente todo o tipo de dado com uma componente espacial pode ser representado como um mapa e sofrer as manipulações e transformações necessárias para a análise que se queira fazer.

O elemento fundamental de informação espacial $T = \langle x, y, z_1, z_2, \dots, z_n \rangle$ onde n representa os valores de variáveis de espaço (z) na localização (x, y) proposto por Goodchild (1992) torna-se realizável quando se interage a capacidade dos GIS e do SR para geração de dados e a capacidade das RNA para trabalhar dados de formatos diferentes e de fontes diversas.

A espacialização de dados alfa-numéricos presente nos bancos de dados censitários de empresas públicas e privadas permite uma nova forma de interpretação e amostragem desses dados.

Berry (1993) indica que procedimentos analíticos não espaciais (i.e. estatística convencional) resultam em tremenda perda de informação que se tem no mapa original. O que se depreende é que os mapas digitais contém detalhes de variação espacial saindo de sua condição meramente analógica para quantificar sistemas físicos, sociais ou econômicos.

O autor argumenta que o impacto total da capacidade de representação numérica dos GIS encontra-se na infância de seu reconhecimento.

Algumas limitações devem ser consideradas quando esse tipo de representação é utilizado.

A escala de trabalho tem influência significativa na definição das variáveis de interesse, especificamente quando se analisa os fatores indutores de mudança na paisagem. A espacialização de dados censitários sociais, biofísicos e econômicos utilizando-se da divisão política pode ser redundante quando se trabalha com escalas maiores que 1:50.000 (1:25.000; 1:2.000).

Miller et al., (1995) comenta sobre esse fato que é contraproducente aumentar o tamanho da rede para incorporar dados que tenham o mesmo valor em qualquer situação.

A utilização dessa forma de informação faz mais sentido quando se trabalha em escalas menores que 1:50.000 (1:100.000; 1:250.000 e 1:1000.000). Com a resolução da imagem de satélite menor (120 x 120m; 1.1 x 1.1 Km) o pixel pode representar mais facilmente um número maior de informação. Isso se dá porque no processo análise a amostragem é feita em áreas maiores englobando um número muito mais significativo de municípios e o efeito que a sua estrutura tem na mudança da paisagem.

A definição dos limites da área a ser analisada tem significado fundamental para interpretação dos fenômenos indutores de mudança. Infelizmente a capacidade de processamento dos computadores pessoais é limitada e esse tipo de análise começa a se tornar corporativo.

Esse fato remete às discussões das muitas definições dos GIS na seção 2.5. Do presente trabalho depreende-se que conforme se amplia a área da paisagem para análise, o grau de dificuldade em se lidar com o problema cresce exponencialmente. Neste ponto cabe concordar com alguns autores que argumentam que os GIS são sistemas altamente corporativos, necessitando de forte infra-estrutura de recursos humanos, financeiros e institucionais.

O SR apontou suas limitações quando se procura obter séries históricas de dados. Entre os vários fatores limitantes podem ser citados o custo elevado das imagens tornando a sua utilização necessariamente via institucional.

Quando se dispõe dessa forma de dados nas instituições outra barreira a enfrentar é o corporativismo das mesmas que impede uma disseminação mais efetiva do conhecimento.

Como agravante, o formato das imagens de satélite no Brasil é não aberto (formato proprietário) e sua leitura e migração pelos sistemas digitais (GIS e SR) é de difícil realização.

Finalmente, as limitações dos softwares disponíveis e suas respectivas versões. O GIS desenvolvido para computadores pessoais (PC) teve um impulso significativo na presente década. Entretanto sistemas operacionais proprietários não contemplam funções que possam converter arquivos de versões menos atualizadas dos programas utilizados.

Em outras palavras, a aquisição e disponibilidade de dados, recursos humanos, máquinas e programas assume definitivamente um aspecto corporativo.

O trabalho de pesquisa individual pode ter repercussão e utilidade como disseminador de conhecimento dentro do universo acadêmico. As aplicações comerciais do GIS, agora aliadas com a IA, têm futuro promissor, somente limitada pela criatividade.

Solução generalizada

Toda a paisagem é influenciada pelo ser humano e sua cultura.

A condução de uma análise econométrica em determinada paisagem pode indicar a propensão que os processos ecológicos, as forças de mercado e os fatores sociais influenciam o uso da terra.

A utilização desses dados econométricos em modelos de simulação de mudanças na paisagem podem ser feitas para se estimar o efeito que determinados usos da terra possam impactar ou não a estrutura e a forma das paisagens.

O impacto das mudanças na integridade do meio ambiente e nos recursos naturais pode ser estimada. Esses impactos podem ser analisados para se avaliar sua influência na sustentabilidade do meio ambiente.

5.2 Sugestões para Futuros Trabalhos

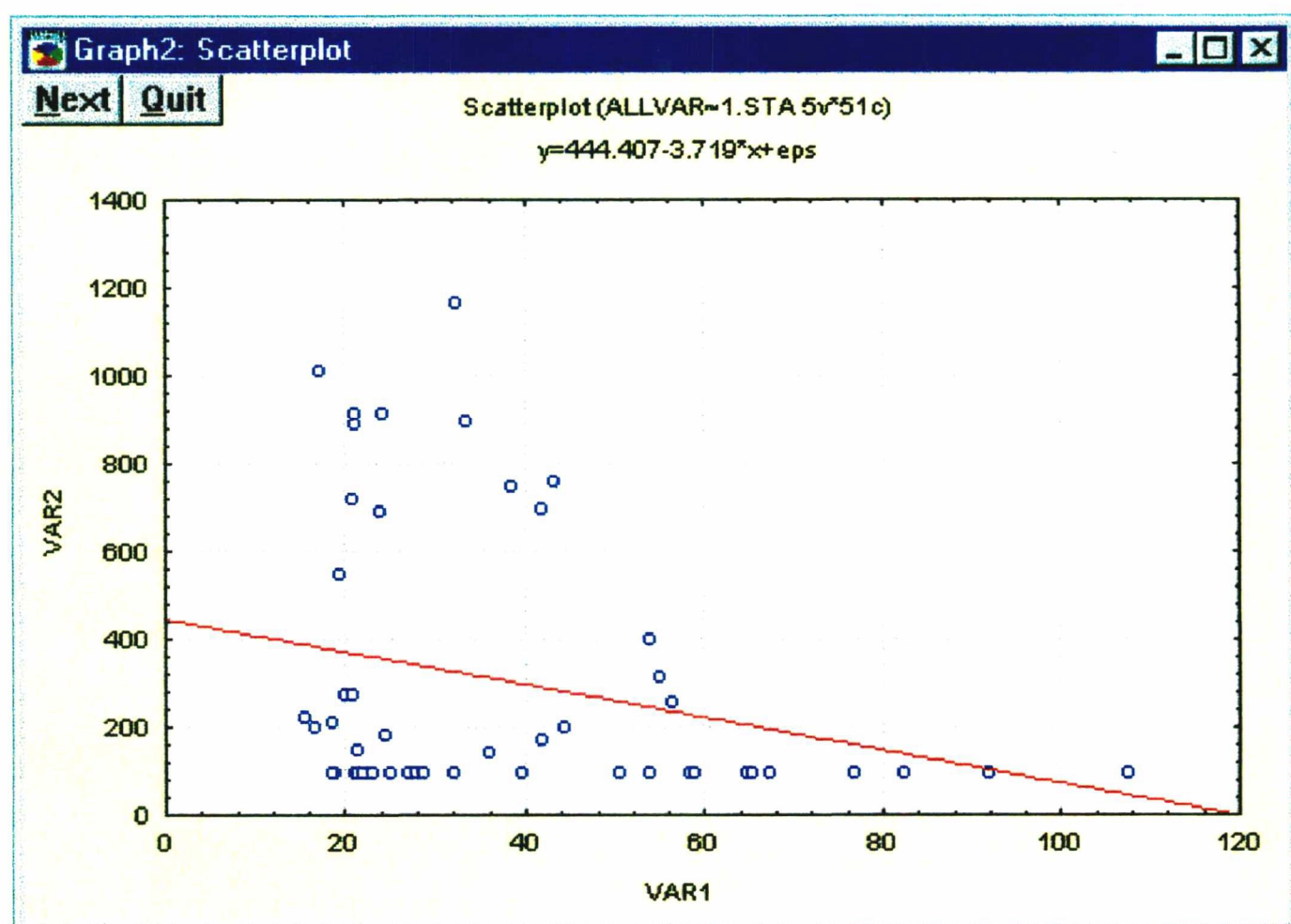
- Através de uma série temporal de imagens de satélite da mesma área de interesse, estimar as probabilidades de transição (matriz de transição) de mudança da cobertura da terra. A partir daí, gerar uma imagem que represente a variável dependente do presente modelo como as probabilidades de mudança que cada pixel individualmente tem;

- Integração de variáveis através da análise do Principal Componente (PCA) e a funcionalidade de sobreposição (“overlay”) em GIS para gerar avaliações mais próximas da realidade;

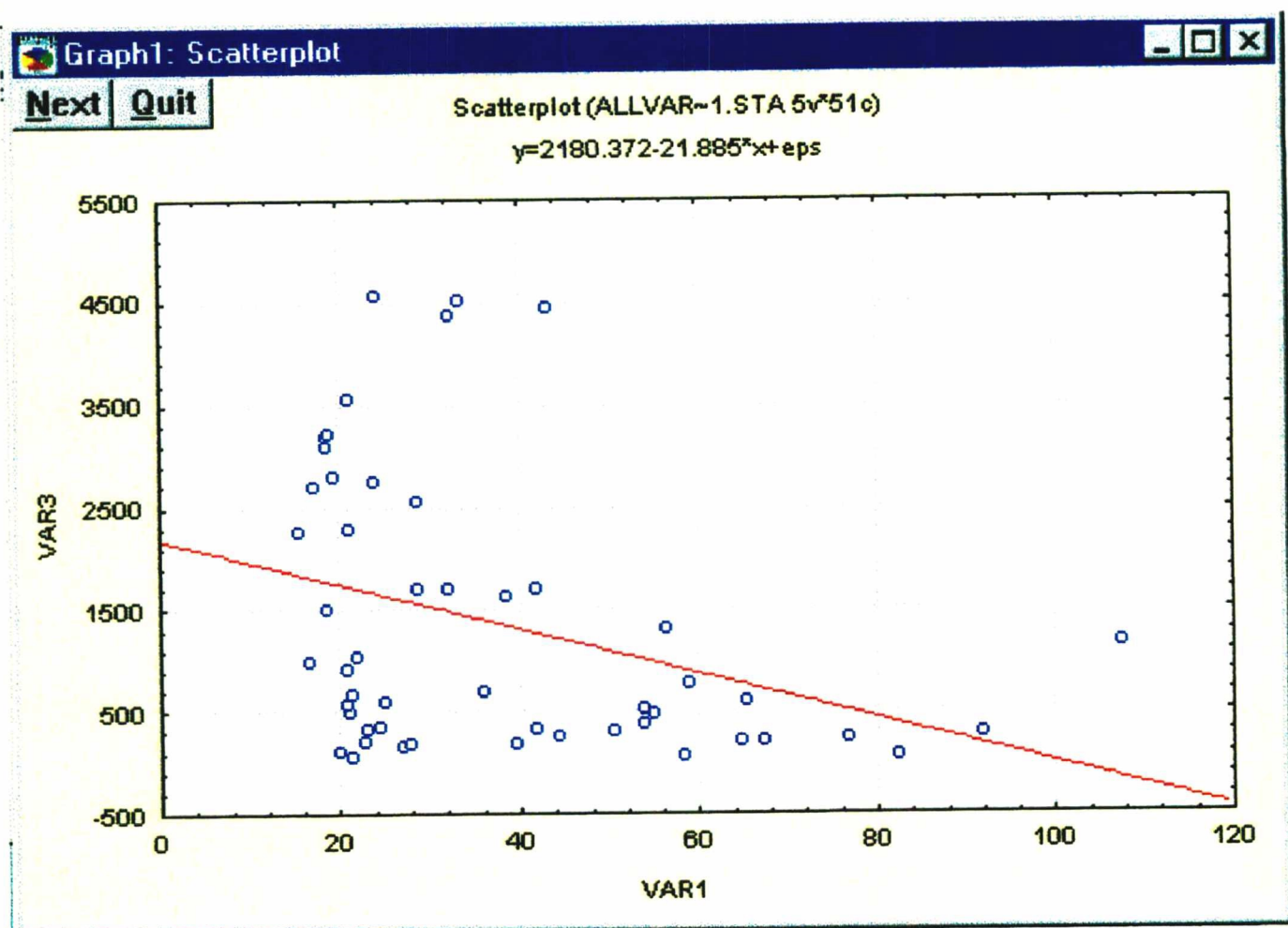
- Utilização do método em escalas menores (1:100.000; 1:250.000; 1:1.000.000) para verificar a significância das variáveis independentes em função de variação de escala;

- Utilização da técnica de RNA “radial-basis” como uma forma de ajustamento de curva (“curve-fitting”) ao invés da aproximação estocástica feita pela “backpropagation”.

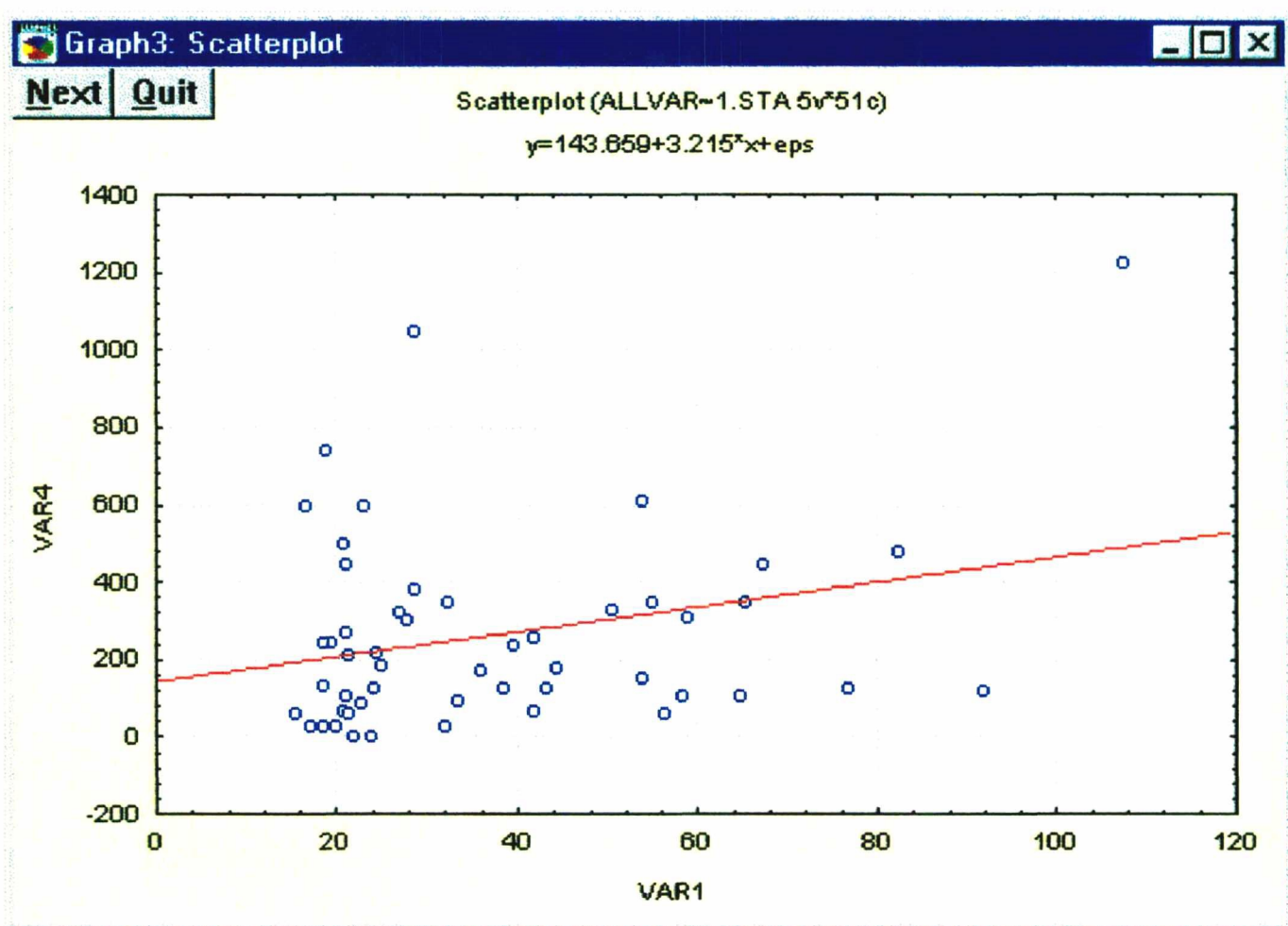
ANEXO 1



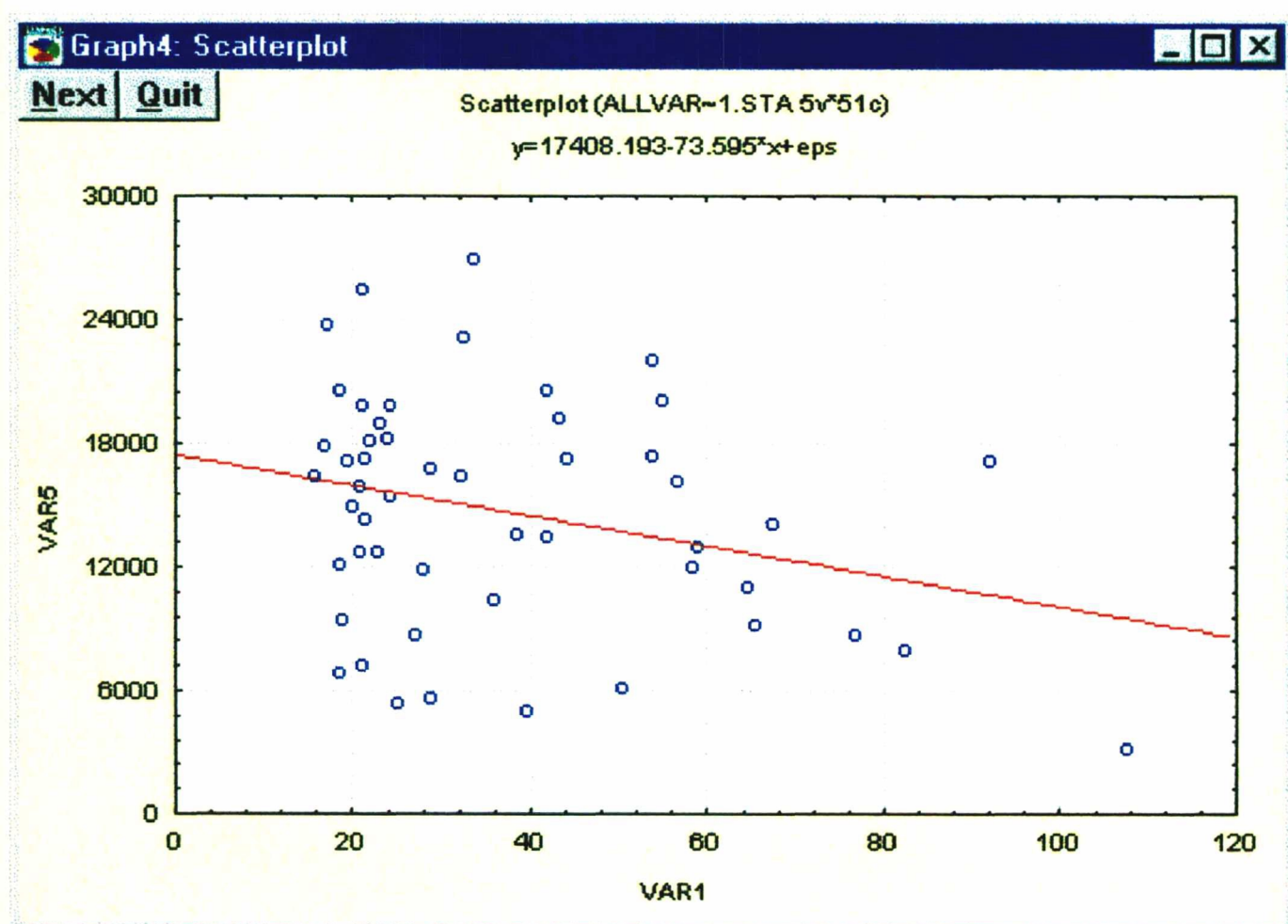
ANEXO 2



ANEXO 3



ANEXO 4



6. Referências

- Anderson, J.R. (1976) A land use and land cover classification system for use with remote sensor data. U.S. Department of Interior, *Geological Survey Professional Paper 964*, 27 pp.
- Anselin, L. (1993) Discrete space autoregressive models. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 454 –469, Oxford University Press, New York.
- Anselin, L., & Getis, A. (1993) Spatial statistical analysis and geographic information systems. In *Geographic Information Systems, Spatial Modeling, and Policy Evaluation*, Manfred M. Fischer and Peter Nijkamp (eds.), pp. 35 –49, Springer-Verlag, Berlin.
- Arnold, U., Datta, B. & Haenscheid, P. (1989) Intelligent geographic information systems (IGIS) and surface water modeling. In *New Directions for Surface Water Modeling (Proceedings of the Baltimore Symposium)* IAHS Publ. No. 181, pp. 407 – 416.
- Baker, & W.L. (1989) A review of models of landscape change. *Landscape Ecology*, 2:111 – 133.
- Baker, W.L., Egbert, S.L. & Frazier, G.F. (1991) A spatial model for studying the effects of climatic change on the structure of landscape subject to large disturbances. *Ecological Modeling*, 56: 109 – 125.
- Batty, P. (1992) Exploiting relational database technology in a gis. *Computers & Geosciences*, 18:453-462.
- Benediktsson, J. A., Swain, & P.H. Ersoy, O. K. (1990) Neural Network Approaches Versus Statistical Methods in Classification of Multisource

Remote Sensing Data. *IEEE Transactions on Geoscience and Remote Sensing*, **28**:540 – 552.

Berry, J.K. (1993) Cartographic modeling: the analytical capabilities of gis. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 454 – 469, Oxford University Press, New York.

Berry, J.K. (1993) Beyond Mapping - Concepts, algorithms and issues in GIS. GIS World Books. Colorado, U.S.A.

Berry, J.K. (1995) Spatial Reasoning for effective GIS. GIS World Books. Colorado, U.S.A.

Binaghi, E., & Rampini, A. (1993) Fuzzy decision making in the classification of multisource remote sensing data. *Optical Engineering*, **32(6)**:1193 – 1204.

Bocco, G. & Valenzuela, C.R. (1993) Integrating satellite-remote sensing and geographic information systems technologies in gully erosion research. *Remote Sensing Reviews*, **7**:233 – 240.

Bolstad, P.V. & Lillesand, T.M. (1992) Rule-based classification models: flexible integration of satellite imagery and thematic spatial data. *Photogrammetric Engineering & Remote Sensing* **58(7)** 965 – 971.

Braun, M.V. (1993) The use of GIS in assessing exposure and remedial alternatives at superfund sites. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.) pp. 339 – 347, Oxford University Press, New York.

- Burrough, .P.A. (1986) Principles of geographic information systems for land resources assessment. *Monographs on Soil and Resources Survey*, 193pp. Claredon Press, Oxford.
- Burrough, P.A. (1989) Matching spatial databases and quantitative models in land resource assessment. *Soil Use and Management*, **5(1)**: 3 – 8.
- Campbell, J.C., Radke, J. Gless, J.T. & Wirtshafter R.M. (1992l) An application of linear programming and geographic information systems: cropland allocation in Antigua. *Environment and Planning*, **24**: 535 – 549.
- Carver, S.J. (1991) Integrating multi-criteria evaluation with geographical information systems. *International Journal Geographical Information Systems*, **5**: 321 – 339.
- Cheng, B & Titterington, D.M. (1994) Neural networks: a review from a statistical perspective. *Statistical Science*, **9(1)**: 2 – 54.
- Cihlar, J., St.-Laurent, L. & Dyer, J.A. (1991) Relation between the normalized difference vegetation index and ecological variables. *Remote Sensing Environment*, **35**: 279 – 298.
- Clothiaux, E. E. & Bachmann, C. M. (1994) Neural Networks and their Applications. In *Neural Nets: Applications in Geography*, Bruce C. Hewitson and Robert G. Crane (eds.), pp. 11 – 21, Kluwer Academic Publishers.
- Constanza, R. (1989) Model Goodness of Fit: A Multiple Resolution Procedure. *Ecological Modeling*, **47**: 199 – 215.
- Cressie, N. & Ver Hoef, J.M. (1993) Spatial statistics analysis of environmental and ecological data. In *Environmental modeling with gis*,

M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 404 – 413, Oxford University Press, New York.

Dale, V.H.,(1994) Causes and effects of land-use change in central Rondônia, Brazil. *Photogrammetric Engineering & Remote Sensing*, **59(6)**: 997 – 1005.

Dale, V.H., Southworth, F. & Pedlowski, M. (1994) Modeling Effects of land Management in the Brazilian Amazonian Settlement of Rondônia. *Conservation Biology*, **8(1)**: 196 – 206.

Demuth, H., & Beale M. (1993) Neural Network TOOLBOX. In *Neural Network TOOLBOX*, The Math Works Inc.

Dobson, J. E. (1993) Commentary: a conceptual framework for integrating remote sensing, gis, and geography. *Photogrammetric Engineering & Remote Sensing*, **59(10)**: 1491 – 1496.

Drayton, R.S., Wilde, B.M. & Harris, J.H.K. (1992) Geographical information system approach to distributed modeling. *Hydrological Processes*, **6**: 361–368.

Dunn, P.C., Sharpe, D.M., Guntenspergen, G.R., Stearns, F. & Yang, Z. (1994) Methods for analyzing temporal changes in landscape pattern. In *Quantitative Methods in Landscape Ecology*, Monica G. Turner & Robert H. Gardner, (eds.), pp. 173 – 198, Springer-Verlag, New York.

Eastman, J.R. (1995) Idrisi for Windows - User's guide. *Clark Labs for Cartographic Technology and Geographic Analysis*, IDRISI Production - Clark University.

- Eastman, J.R., McKendry, J.E. & Fulk, M.A. (1995) Change and Time Series Analysis. *Explorations in Geographic Information Systems Technology* United Nations Institute for Training and Research. Geneva, Switzerland.
- Eddy, J.A. (1993) Environmental research: what we must do. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 3 – 7, Oxford University Press, New York.
- Englund, E.J. (1993) Spatial simulation: Environmental Applications. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 432 – 437, Oxford University Press, New York.
- ESRI (1995) Understanding gis - the ARC/INFO method. *Environmental Systems Research Institute, Inc.* 406pp.
- FAO (1976) A Framewok for Land Evaluation. Food and Agriculture organization of the United Nations. *FAO Soils Bulletin* 32, 72pp.
- Fedra, K. (1993) GIS and Environmental Modeling. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 36 – 50, Oxford University Press, New York.
- Flamm, R.O. & Turner, M.G. (1992) GIS applications perspective: multidisciplinary modeling and gis for landscape management. In *Remote Sensing anf GIS in Ecosystem Management*, **13**: 201 – 212.
- Flamm, R.O. & Turner M.G. (1994) Alternative model formulations for a stochastic simulation of landscape change. *Landscape Ecology* **9(1)**: 37–46.
- Foody, G.M., McCulloch, M.B. & Yates, W.B. (1995) Classification of Remotely Sensed Data by an Artificial Neural Network: Issues Related to

Training Data Characteristics. *Photogrammetric Engineering & Remote Sensing* **61(4)**: 391– 401.

Forman, R.T.T. & Godron, M. (1986) *Landscape Ecology*. 619pp. John Wiley & Sons, New York.

Forman, R.T.T. (1995) Some general principles of landscape and regional ecology. *Landscape Ecology* **10(3)**:133 – 142.

Frank, A.U. (1992) Spatial concepts, geometric data models; and geometric data structures. *Computers & Geosciences* **18**: 409 – 417.

Fung, T. (1987) Application of principal component analysis to change detection. *Photogrammetric Engineering & Remote Sensing* **53(12)**: 1649 – 1658.

Gardner, R.H., Milne, B.T., Turner, M.G., O'Neill, R.V. (1987) Neutral models for the analysis of broad-scale landscape pattern. *Landscape Ecology* **1(1)**:19 – 28.

Goodchild, M.F. (1992) Geographic data modeling. *Computers & Geosciences* **18**: 401– 408.

Goodchild, M.F. & Shiren Y. (1992) A hierarchical spatial data structure for global geographic information systems. *Graphical Models and Image Processing*, **54(1)**: 31 – 44.

Goodchild, M.F. (1993) The state of gis for environmental problem-solving. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 9 – 15, Oxford University Press, New York.

- Gould, P., Kabel, J., Gorr, W. & Golub, A. (1991) Aids:predicting the next map. *INTERFACES*, **21(3)**: 80 – 92.
- Hall, G.B., Wang, F. & Subaryono (1992) Comparison of boolean and fuzzy classification methods in land suitability analysis by using geographical information systems. *Environment and Planning*, **24**:497 – 516.
- Harris, P.M. (1995) The integration of geographic data with remotely sensed imagery to improve classificatin in an urban area. *Photogrammetric Engineering & Remote Sensing* **61(8)**: 993 – 998.
- Haykin, S. (1994) NEURAL NETWORKS A Comprehensive Foundation. 696pp. Macmillan College Publishing Company, Inc. New Jersey.
- Hewitson, B.C. & Crane, R.G. (1992) Large-scale atmospheric controls on local precipitation in tropical Mexico. *Geophysical Research Letters*, **19(18)**: 1835 – 1838.
- Hewitson, B.C. (1994) Precipitation controls in southern mexico. In *Neural Nets: Applications in Geography*, Bruce C. Hewitson and Robert G. Crane (eds.), pp. 121 – 143, Kluwer Academic Publishers, Dordrecht.
- Hewitson, B.C. & Crane, R.G. (1994) Looks and Uses. In *Neural Nets: Applications in Geography*, Bruce C. Hewitson and Robert G. Crane (eds.), pp. 1 – 8, Kluwer Academic Publishers, Dordrecht.
- Hofstadter, D. (1994) Fluid Concepts and Creative Analogies. *Computer Models of the Fundamental Mechanisms of Thought*. 518pp. HarperCollins Publishers, Inc. New York.

- Hooper, R.P. & Peters N.E. (1989) Use of multivariate analysis for determining sources of solutes found in wet atmospheric deposition in the United States. *Environmental Science Technology*, **23**: 1263 – 1268
- Hulshoff, R.M. (1995) Landscape indices describing a Dutch landscape. *Landscape Ecology* **10**(2):101-111.
- Hunsaker, C.T., Nisbet, R.A., Lam, D.C., Browder, J.A., Baker, W.L., Turner, M.G. & Botkin, D.B. (1993) Spatial models of ecological systems and processes: the role of GIS. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 248 – 264, Oxford University Press, New York.
- Hutchinson, C.F. (1982) Techniques for combining landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering & Remote Sensing* **48**(1): 123 – 130.
- Johnston, R.J. (1978) *Multivariate Statistical Analysis in Geography*. 280 pp., Longman, London.
- Johnston, C. A. (1993) Introduction to quantitative methods and modeling in community, population, and landscape ecology. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 276–283, Oxford University Press, New York.
- Janssen, R. & Rietveld, P. (1990) Multicriteria analysis and geographical information systems: an application to agricultural land use in the Netherlands. *Geographical Information Systems for Urban and Regional Planning*, pp. 129-139.

- Kemp, K.K. (1990) Spatial databases: sources and issues. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 363– 371, Oxford University Press, New York.
- Lam, D.C.L. (1993) Combining ecological modeling, gis and expert systems: a case study of regional fish species richness model. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 270– 275, Oxford University Press, New York.
- Lee, R.G., Flamm, R. & Turner, M.G. (1992) Integrating sustainable development and environmental vitality: a landscape ecology approach. In *Watershed Management: Balancing Sustainability and Environmental Change*, Robert J. Naiman (ed.), pp. 499-521, Springer-Verlag, New York.
- Lee, C.T. (1995) The use of archival landsat mss and ancillary data in a gis environment to map historical change in a urban riparian habitat. *Photogrammetric Engineering & Remote Sensing* **61(8)**: 999 – 1008.
- Lillesand, T.M. & R.W. Kiefer, 1987. Remote Sensing and Image Interpretation, John Wiley and Sons, Nova York.
- Lo, C.P. (1996) Integration of landsat thematic mapper (TM) data and U.S. census data for quality of life assessment. *International Archives of Photogrammetry and Remote Sensing*, vol. XXXI, part B7, Viena.
- Lowell, K.E. (1993) Predictive model development and evaluation with unknown spatial units. *Photogrammetric Engineering & Remote Sensing*, **59(10)**: 1509 – 1515.
- Malila., W.A. (1980) Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat. *Proceedings, 6th Annual Symposium on Machine Processing of Remotely Sensed Data*, pp.326-335.

- McGinnis, D.L. (1994) Predicting snowfall from synoptic circulation: a comparison of linear regression and neural network methodologies, In *Neural Nets: Applications in Geography*, Bruce C. Hewitson and Robert G. Crane (eds.), pp. 79 – 99, Kluwer Academic Publishers, Dordrecht.
- Meentemeyer, V. (1989) Geographical perspectives of space, time, and scale. *Landscape Ecology*, **3(3/4)**:163 – 173.
- Michalek, J.L., Luczkovich, J.J. & Stoffle, R.W. (1993) Multispectral change vector analysis for monitoring coastal marine environments. *Photogrammetric Engineering & Remote Sensing*, **59(3)**: 381 – 384.
- Miller, D. M., Kaminsky, E.J. & Rana, S. (1995) Network Classification of Remote Sensing Data. *Computers & Geosciences*, **21(3)**: 377 – 386.
- Nyerges, T.L. (1993) Understanding the scope of gis: its relationship to environmental modeling. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 75 – 93, Oxford University Press, New York.
- Nyerges, T.L. (1989) Schema integration analysis for the development of GIS databases. *International Journal Geographical Information Systems*, **3**: 153 – 183.
- Nyerges, T.L. (1991) Geographic information abstractions; conceptual clarity for geographic modeling. :1483 – 1499.
- Odland, J. (1988), Spatial autocorrelation, 85pp., SAGE Publications, California.

O'Neill, R.V., Krummel, J.R., Gardner, R.H., Sugihara, G., Jackson, B., De Angelis, D.L., Milne, B.T., Turner, M.G., Zygmunt, B., Christensen, & S. (1988) Indices of landscape patterns. *Landscape Ecology*, **1(3)**: 153 – 162.

Openshaw, S. (1993) Modeling spatial interaction using a neural net. In *Geographic Information Systems, Spatial Modeling, and Policy Evaluation*, Manfred M. Fischer and Peter Nijkamp (eds.), pp. 147 – 164, Springer-Verlag, Berlin.

Openshaw, S. (1993) Some suggestions concerning the development of artificial intelligence tools for spatial modeling and analysis in GIS. In *Geographic Information Systems, Spatial Modeling, and Policy Evaluation*, Manfred M. Fischer and Peter Nijkamp, (eds.), pp. 17 – 33, Springer-Verlag, Berlin.

Openshaw, S. (1994) A concepts-rich approach to spatial analysis, theory generation, and scientific discovery in GIS using massive parallel computing. In *Innovations in GIS - selected papers from the First National Conference on GIS Research UK*, Michael F. Worboys, (ed.), pp. 123 – 137, Taylors & Francis, UK.

Openshaw, S. (1994a) Neuroclassification of Spatial Data. In *Neural Nets: Applications in Geography*, Bruce C. Hewitson and Robert G. Crane (eds.), pp. 53 – 68, Kluwer Academic Publishers, Dordrecht.

Pao, Y. (1989) Adaptive Pattern Recognition and Neural Networks. In , pp. Addison-Wesley Publishing Company, Inc. Reading.

Quattrochi, D.A. & Pelletier, R.E. (1994) Remote Sensing for Analysis of Landscapes: an introduction. In *Quantitative Methods in Landscape*

Ecology, R. Haines-Young, David R. Green and Stephen H. Cousins (eds.), 3: 51 – 73. Taylor & Francis, New York.

Richards, J.A. (1993) Remote sensing digital image analysis - an introduction. 340 pp., Springer-Verlag, New York.

Ritter, N.D., Logan, T.L. & Bryant, N.A. (1994) Integration of neural network technologies with geographic information systems. *Jet Propulsion Laboratory, California Institute of Technology*, Pasadena, CA.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986) Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the microstructures of Cognition*, D.E. Rumelhart and J.L. McClelland (eds.), vol. 1, pp. 318 – 362, MIT Press, Cambridge, MA.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986) Learning representations by back-propagating errors. *Nature*, 323: 533 – 536.

Rifai, H.S., Newell, C.J. & Bedient, P.B. (1993) GIS enhances water quality modeling. *GIS WORLD*, august, pp. 52 – 55.

Silveira, V.F., Khator, S.K. & Barcia, R.M. (1996) An Information Management System for forecasting environmental change. 19th *International Conference on Computers and Industrial Engineering*. Miami – Florida

Schaller, J. (1990) Geographical Information System Applications in Environmental Impact Assessment. In *Geographical Information Systems for Urban and Regional Planning*, H.J Scholten and J.C.H. Stillwell (eds.), pp. 107 – 117, Kluwer Academic Publishers. The Netherlands.

Smith, T.R. (1984) Artificial intelligence and its applicability to geographical problem solving. *The Professional Geographer*, 36(2): 147 – 158.

- Smith, M.B. & Vidmar, A. (1994) Data set derivation for gis-based urban hydrological modeling. *Photogrammetric Engineering & Remote Sensing* **60(1)**:67-76.
- Singh, A. (1989) Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sensing*, **10(6)**: 989 – 1003.
- Smith, M. (1994) Neural Networks for Statistical Modeling. Van Nostrand Reinhold. 178pp. New York.
- Star, J. & Estes, J. (1990) Geographic information systems - an introduction. 303pp. Prentice-Hall Inc., New Jersey.
- Steyaert, L.T. (1993) Spatial data. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 361 – 362, Oxford University Press, New York.
- Swain, P.H., Richards, J.A. & Lee, T. (1985) Multisource data analysis in remote sensing and geographic information processing. *Machine Processing of Remotely Sensed Data Symposium*, pp.211 – 218.
- Tomlin, C.D. (1990) Geographic information systems and cartographic modeling. 249pp. Prentice-Hall Inc., New Jersey.
- Turner, S.J., O'Neil, R.V., Conley, W., Conley, M.R. & Humpries, H.C. (1994) Pattern and Scale: Statistics for Landscape Ecology. In *Quantitative Methods in Landscape Ecology*, R. Haines-Young, David R. Green and Stephen H. Cousins (eds.), **2**:17 – 49. Taylor & Francis, New York.
- Turner, M.G. & Gardner, R.H. (1994) Quantitative methods in landscape ecology: an introduction. In *Quantitative Methods in Landscape Ecology*,

R. Haines-Young, David R. Green and Stephen H. Cousins (eds.), 1:3 – 14. Taylor & Francis, New York.

Turner, M.G. (1987) Spatial simulation of landscape change in Georgia: a comparison of three transition models. *Landscape Ecology* 1:29 – 36.

Turner, M.G., & Romme, W.H. (1994) Landscape dynamics in crown fire ecosystems. *Landscape Ecology*, 9(1): 59 – 77.

Valenzuela, C.R. (1991) Basic Principles of Geographic Information Systems. In *Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries*, A.S. Belward and C.R. Valenzuela (eds.), pp. 279 – 295. ECSC, EEC, EAEC, Brussels and Luxembourg.

Valenzuela, C.R. (1991:Book) Data analysis and modeling. In *Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries*, A.S. Belward and C.R. Valenzuela (eds.), pp. 335 – 348. ECSC, EEC, EAEC, Brussels and Luxembourg.

Valenzuela, C.R. (1991:Book) Spatial Databases. In *Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries*, A.S. Belward and C.R. Valenzuela (eds.), pp. 311 – 333.. ECSC, EEC, EAEC, Brussels and Luxembourg.

Valenzuela, C.R., H. Brouwer, & A. Meijerink (1991:Book) Land use model using a geographic information system. In *Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries*, Belward AS, Valenzuela CR (eds.), pp. 425 – 441, ECSC, EEC, EAEC Brussels and Luxemburg.

- Yamamoto, Y. (1996) The applications of neural networks to GIS in the construction of land evaluation models. *International Archives of Photogrammetry and Remote Sensing*, vol. XXXI, part B7, pp. 788 – 793 Vienna.
- Wadge, G., Wislocki, A.P. & Pearson, E.J. (1993) Spatial analysis in GIS for natural hazard assessment. In *Environmental modeling with gis*, M.F. Goodchild, B. O. Parks and L. T. Steyaert (eds.), pp. 332 – 338, Oxford University Press, New York.
- White, H. (1989) Learning in artificial neural networks: a statistical perspective. *Neural Computation*, 1: 425 – 464.
- Widrow, B. Hoff, M.E, & . (1960) Adaptive switching circuits. IRE WESCON Convention Record New York pp. 96 –104.
- Winter, K. & Hewitson (1994) Self organizing maps -- application to census data. In *Neural Nets: Applications in Geography*, Bruce C. Hewitson and Robert G. Crane (eds.), pp. 71 – 77, Kluwer Academic Publishers, Dordrecht.
- Zhou, J. & Civco, D.L. (1996) Using genetic learning neural networks for spatial decision making in gis. *Photogrammetric Engineering & Remote Sensing* 62(11):1287 – 1295.